

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2001, 03:32:18 ; Search time 48.09 Seconds  
(without alignments)  
599.061 Million cell updates/sec

Title: US-09-446-677B-8

Sequence: 1 MKIPURFLILSLVPTLSMSN.....FEMGRSSRYNVDAGSKIKF 841

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	525.5	12.2	964	1	OMPE_CHLTR
2	360.5	8.4	1034	1	OMPE_CHLTR
3	235	5.5	1250	1	YFAL_ECOLI
4	212	4.9	1645	1	OMPR_RICTY
5	204	4.7	1041	1	EGRT_YEAST
6	203.5	4.7	1654	1	OMPR_RICRI
7	199.5	4.6	2003	1	YDBA_ECOLI
8	197	4.6	933	1	SLAP_CAME
9	193	4.5	1300	1	120K_RICRI
10	188.5	4.4	1569	1	YPRJ_ECOLI
11	187	4.4	2249	1	190K_RICRI
12	186	4.3	1643	1	OMPR_RICPR
13	184	4.3	1656	1	OMPR_RICJA
14	182	4.2	1039	1	AG43_ECOLI
15	178.5	4.2	1025	1	SLAP_CAVCR
16	172.5	4.0	1291	1	VAC4_HELPY
17	171.5	4.0	1286	1	AIDA_ECOLI
18	171	4.0	1290	1	VACA_HELPY
19	171	4.0	3591	1	FHAB_BORPE
20	169	3.9	995	1	YIO9_YEAST
21	167.5	3.9	574	1	FLB3_CAMJE
22	167	3.9	1577	1	HLVA_PROMI
23	167	3.9	2329	1	YS89_CAEEL
24	165.5	3.9	1287	1	VAC2_HELPY
25	164.5	3.8	881	1	YJH8_YEAST
26	162.5	3.8	574	1	FLA3_CAMJE
27	161.5	3.8	1296	1	VACI_HELPY
28	161	3.7	749	1	TROP_HUMAN
29	160	3.7	575	1	FLB2_CAMJE
30	158.5	3.7	1150	1	APMO_PIG
31	158	3.7	1120	1	STPR_ECOLI
32	157.5	3.7	572	1	FLAB_CAMCO
33	157.5	3.7	587	1	FTBP_ADE12

34	155.5	3.6	1076	1	NUPI_YEAST
35	154	3.6	1608	1	HLVA_SERMA
36	153	3.6	575	1	FLA3_CAMJE
37	153	3.6	1310	1	VAC3_HELPY
38	153	3.6	1567	1	ICEN_XANTC
39	152.5	3.5	572	1	FLA3_CAMCO
40	152.5	3.5	928	1	HXA2_HAEIN
41	150	3.5	1036	1	HP12_DEIRA
42	150	3.5	1200	1	ICEN_PSESY
43	150	3.5	1210	1	ICEN_PSEFL
44	149	3.5	512	1	FLA3_HELPY
45	149	3.5	571	1	FLB1_CAMJE

## ALIGNMENTS

RESULT	1	STANDARD	PRT	964 AA.
OMPE_CHLTR				
ID	OMPE_CHLTR			
AC	084877;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	PUTATIVE OUTER MEMBRANE PROTEIN E PRECURSOR.			
GN	PMPE OR CT869.			
OS	Chlamydia trachomatis.			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI_Taxid=813;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATIN-D/UM-3/CX;			
RX	MEDLINE-9900809; PubMed-9784136;			
RA	Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,			
RA	Mitchell W.P., Ollinger L., Tatusov R.L., Zhao Q., Koonin E.V.,			
RA	Davis R.W.;			
RT	"Genome sequence of an obligate intracellular pathogen of humans:			
RT	Chlamydia trachomatis."			
RL	Science 282:754-759(1998).			
CC	- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)			
CC	(POTENTIAL).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL: AE001360; AAC68467.1; -			
KW	Outer membrane; Signal.			
FT	SIGNAL			
FT	CHAIN 19 964			
FT	SEQUENCE 964 AA; 104703 MW; 1B998A7D2571CE2 CRC64;			

Query Match 12.2%; Score 525.5; DB 1; Length 964;  
Best Local Similarity 23.8%; Pred. No. 5.5e-23;  
Matches 209; Conservative 120; Mismatches 339; Indels 211; Gaps 33;

QY	23	GAATTELSANSNFDGTTTSSTSSATDGTN-----YFKSVIENPKT 72
DB	153	GATIAQMLYINHNHDVGFPMKFNFSYVGGAISTANTFEVSENOSCFPMNICIQT 208
QY	73	GEMTSTSCFNDAAGDLNFGGFTFSNIDATGASGAIGSEANKVTYISGFSALS 132
DB	209	-----NTAKGGAITAGTNSNFSNNCDLFTNNACAGAG----- 244
QY	133	FLASPASTVINGIAINVKGNLSILDND--KVLIDNFSTGCGAICAGSLKIANKS- 189
DB	245	-IFSPICSLTGN-----RGNIVFYNNRCKRVETASSEASDGAIKVTTRLDVTGNRGR 297





```

Db 681 TTPSNTTASIPNIKASGSSGSASNSGEVITFKHTLVNMAPGYIVDPIRGLDIA 740
QY 576 STMPGDDTLISGFADLPARDKYF-----MTNPAK-----TYAGSLRQHDASLSYV 625
Db 741 NSLVHSGRMNMTGLRSLRP-DNSWFLQGAATLFTKOOKRLSYHG---YSSASKGYTAS 796
QY 626 SILLEGGLREILLFPVYSKTLPCSFYGOLSYGTHDRMKTESLPPEPTLSTGHTSGWY 685
Db 797 SQASGANG-HKFLISFSQSS-----DKMEKETNNRLSRYT---LSALCFEHPMFDR1 846
QY 686 WMAGELGTAVAVEN-----TSGRGFERE-----YTPPEVKYQ 716
Db 847 ALIGAACVNGTNNMRSFYCTKSSKGFSTLTGASLRCELNDSPMLRSIMLTPF--AQ 904
QY 717 AVYSROD--SFVGLAISRDFSDSHLYNLA1-PLGIK---LEKRPADQYHYVAVMTSPDV 770
Db 905 ALFSREPASIRSGDLARLFTLEQAHAVASDPDGIGKAVSSDTPWPLSMEMELAOPPL 964
QY 771 CRSNPCTTTLTNSGSMKTKGSNLARQAGIVQASGPRSIG-AAELFGMFGEMGSSSR 829
Db 965 YMKRPLNTLILNNGSWTNTTPLAKHS--FYGRSHSLKESHKLAFANYQAEVATSTV 1022
QY 830 SYNVDAGSKIFE 841
Db 1023 SHYINAGALVF 1034

RESULT 3
YFAL.ECOLI
ID YFAL.ECOLI STANDARD; PRT: 1250 AA.
AC P45508; P45507; P39441; P76468; P77467;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 131.2 KDA PROTEIN IN UBIG-NRDA INTERGENIC REGION
GN PRECORSOR.
DE YFAL.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [12]
RN SEQUENCE FROM N.A.
RP STRAIN-K12;
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono K., Itoh T., Kanai K., Kasai H.,
RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizouchi K., Mori H., Motomura K., Nakamura Y.,
RA Nishino H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RN PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
RX MEDLINE=84272624; PubMed=6087316;
RA Carlson J., Fuchs J.A., Messing J.;
RT "Primary structure of the Escherichia coli ribonucleoside diphosphate
RT reductase operon."
RL Proc. Natl. Acad. Sci. U.S.A. 81:4294-4297(1984).
RN [4]
RN SEQUENCE OF 925-1198 FROM N.A.
RC STRAIN-K12 / EMG2;
RA Estep P., O'Keefe T., Robison K., Church G.M.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

```

```

RN [15]
RP SEQUENCE OF 1180-1250 FROM N.A.
RC STRAIN-OV6;
RX MEDLINE=88201664; PubMed=2834621;
RA Hussain K., Elliott E.J., Salmond G.P.C.;
RT "The paired-mutant of Escherichia coli also carries a gyrA mutation.
RT The complete sequence of gyrA."
RL Mol. Microbiol. 1:259-273(1987).
RN [6]
RP IDENTIFICATION.
RX MEDLINE=96032851; PubMed=7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes."
RL Nucleic Acids Res. 23:3554-3562(1995).
CC -1- SIMILARITY: TO E.COLI YDEK.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
CC MANY FRAMESHIFTS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sid.ch).
CC
DR EMBL; AEO00313; AAC75293.1; -
DR EMBL; D90855; BAA16052.1; ALT_INIT.
DR EMBL; D90854; BAA16050.1; ALT_INIT.
DR EMBL; K02672; -; NOT_ANNOTATED_CDS.
DR EMBL; U30459; AAA74094.1; -
DR EMBL; Y00544; -; NOT_ANNOTATED_CDS.
DR EcoGene; E012850; yfal.
KW Hypothetical protein; Repeat; Signal.
FT SIGNAL 1 23
FT CHAIN 24 1250
FT DOMAIN 919 948
FT CONFLICT 28 30
FT CONFLICT 40 40
FT CONFLICT 65 66
FT CONFLICT 431 431
FT CONFLICT 433 433
FT CONFLICT 478 478
FT CONFLICT 773 773
FT CONFLICT 853 853
FT CONFLICT 923 924
FT CONFLICT 948 994
SQ SEQUENCE 1250 AA; 131152 MW; 17F98C05E299FC93 CRC64;
LRIIGG -> LTRSC (IN REF. 4).

```

Query Match 5.5%; Score 235; DB 1; Length 1250;

Best local Similarity 22.5%; Pred. No. 4,7e-06; Mismatches 268; Indels 224; Gaps 38;

Matches 174; Conservative 109; Mismatches 268; Indels 224; Gaps 38;

```

QY 8 LLSLVPTLSMNLGAAATE-----ELAS-----NSFDGTTSTS-----FSSKTS 50
Db 13 LTPSMIASLFSANGVAAVTSCGYDVKASCQASROSLSGITODWSTADSCOWLVFSDMTN 72
QY 51 SATDGTNYVFKD---SVYINVPKGTGTOSTCPKNAAGDINFGGGSFTFNSNDAT 107
Db 73 MASGAVFLOOGAEFSLIPN--ETGNT---LEANNVTYGETYNGGATFAKENSTLNT 126
QY 108 T-----ASGAIGSEAAKNT--VTLGFSFASLPKSPASTVTVG--LGAIVNKGLS 155
Db 127 DYFSGVAVAGYGGAIVSSGTFNDTGAVDLVTVNAM-FRNNIANDGKGALYTTINNDYLS 185
QY 156 --LLDNDKVLIDQNFSTGGGAINCA-----GSLKANKKSLFTGSSSTRGAT 204
Db 186 DYVFDNNQAAVTSYSDPDGGAIDVTDNNSDKRPSGTYIVNN--TAFNTNNTAGYGAI 243

```

```

OY 205 HFKNLT-----ISSGFTLFGNTAPPAAGKGAIALADSGTSLISGSGDII 252
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 244 YTSVAPYLIDISVSDYSQNGVLVDENNSAAGYGGDSSAAGFMYLGLSEVTFDI 303
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 253 FRENTEGATGVSHASLIDGTSKITALRAAQCHTITFYDPIVNTGSTAADLNTNSPD 312
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 304 -DCKTLVIGTENDGAVD---STAGTGL-----LTKTSGDL-VLN----- 339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 313 TGDNRKCTGVTSYSGEKL-----EAEAKDEKNRTSKL-----LQNV----- 350
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 340 -ADNDFGTGMQJENGEVTLGRSNSLMNVGDTHCQDDPOCCVGLTIGSIDQYQNAELNV 398
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 351 -----FKNGTVLK--GDVYLSANGFS--QDANSKLIMDGTSLVANTEST 392
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 399 GSTQTFVHALTGFGQNTLIDAGGVNTVNOGSFAGILEGAGQLTIAQNSYVLAQAQSM 458
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 393 ELTNLEINIDSLNRGKKIKLSAATAQKDIRIDRPV-----VLASIDSEFYONGFLNED 445
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 459 ALTG-DIVVD--DGAVLSEGDADULTALQDDPOSITVLNGVLDLSDFTWOSG----- 509
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 446 HSYDGLLELDA-----GKD---IVISADSRINAQSPYQ 477
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 510 TSTNDLELVSGSGTYIGSDVVDLAGDNLHTGDKGKGVYVVDASDQVSLANN-NS 568
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 478 YQG-----KWTINWSTDCKKATVSMAKQSENPTEQEPAPLVNLLGSEFIDVRFQNFIE 532
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 569 YLGTQTIASGTLVNSDQSLGDTNRYNRQVIFTDKQES-----VMEITSVDV 615
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 533 LGTEGAPYKRFVNAVATSNVLRHSGRENQKRFHVSGGAVVNGAST--RMDGGTSLIGRA 590
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 616 TTSDAAGHRDIEMRADGEVAVDAGVDT-----WGLAMADSSOHODEGSTLT----- 664
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 591 QLFARDKDFMTNTPFAKTYASLRLQHDASLYSVSILLG--EGGRREILLPYVS 643
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 665 -----KTGAGTLELTAAGTGTQSAVAVEGTLKGDVADI-LPYAS 702
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RA Hackstadt T., Messer R., Cieplik W., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent
RT mutant deficient in processing.";
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04661; AAB48987.1; -.
KW Antigen; S-layer; Transmembrane; Cell wall.
FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1 1354 1645 32 KDA BETA PEPTIDE.
FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
FT CONFLICT 657 657 H -> N (IN REF. 2).
FT CONFLICT 842 842 V -> I (IN REF. 2).
FT CONFLICT 1071 1071 G -> A (IN REF. 2).
FT CONFLICT 1306 1306 G -> S (IN REF. 2).
SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

Query Match 4.9%; Score 212; DB 1; Length 1645;
Best Local Similarity 21.0%; Pred. No. 0.00015;
Matches 216; Conservative 119; Mismatches 310; Indels 384; Gaps 55;

```

```

RESULT 4
ID OMPB_RICTY STRAND: PRT; 1645 AA.
AC P96989;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN SLP
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WILMINGTON;
RA MEDLINE=94040787; PubMed=8224886;
RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT "Cloning and sequence analysis of the gene encoding the crystalline
RT surface layer protein of Rickettsia typhi.";
RL Gene 133:129-133(1993).
RN [2]
RP PARTIAL SEQUENCE.
RC STRAIN=WILMINGTON;
RA MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.";
RL Mol. Immunol. 29:95-105(1992).
RN [3]
RP IDENTIFICATION OF CLEAVAGE SITE.
RA MEDLINE=92104668; PubMed=1729180;

```

```

OY 26 TTBELASNSFDGTTSTFSKTSATDGTN---YFEDSVYIE-NVKT--GEFQSTS 79
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 246 TTPDANANLLOGGGNTINFNGR---DGTGLVLSKNGNTEPNTVNTIKAEVE 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 80 CFKNDAAAGDLNPLFG-----GFSTFENIDATTSAGAIQSEANKTV 123
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 301 EBDTAAAGKLIANGAANAIVCTDNGACRAGFIYSVNGNANATISQVYAKDIYQSA 360
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 124 TUSGFALSFLKSPASTVINGALAVKGNLSLLDNKVLIDQNESTGDCGAINCAGSIK 183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 361 NAGGQYTFEHL-----YVVGIGG--KTNFTADS-KVILITENASFCSTDPGNLAVQIV 410
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 184 IANKSL--SFTG--NSSSTRGAIHTKNTLTSGGFTLPGCTAPTA----- 228
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 411 VPNNKILTSNFIQDAKNNKNTAG-----VITFNANG-TLVSGNTDPNIVTNIKAEVE 463
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 229 -----GKGAIAIADSGTSLISGD-----SGDIIEGNTI 258
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 464 GAGIYVLSGIRHAEELRLGNAGSIFKLADGV--INGVYNNPPLVNNALAGSIQDGSNI 522
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 259 GATGVSHSAID-----LGTSAKITLARAQ-----GHTIYF----- 290
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 523 -ITGDIGNCAVNAALQDITFLANDASKILTLGSAITIGANAGAIHFQANGCTIQLTSTON 581
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 291 -----YDPTTVGTSVADA-----LNINSPDT-----G 314
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 582 NILVDFDLDTTDDGTGVVDASSLTNNQTLTNGSIGTIGANTKLGIRFVNGSSKTLILNAG 641
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 315 D-----NKEYTGVTFSGEKL-TEAEAKDEKNRTS--K 344
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 642 DVAINELVWENDGSVHLTNTYLIKTINAANOQKIIVAADPIINDTALADGTNLGSAIS 701
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 345 LQNVAF-----KNGTVVL--KGDVYLSANGFSODAN--SKLIMDGTSLVANT-----ES 391
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 702 PLSNIHFATKANQDSILHIGKGVNLYANNITTTDANVCSLHFRSGGTSIVSGYGGQGG 761
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 392 IELTNLEINIDSLNRGKKIK-LSAATAQKDIRIDRPVLAISD-----ESFYONG--- 440
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 762 LKLNLT-----IIDNGTVAFFIDIFNGSGTKIEGKSILLOISSNYITDIESADNCTLE 816
QY 441 -----FLNEDHSYDGLLELDACKDIYISADSRISINAVOSPYGKWTJNMSTDCK 491
Db 817 FVNTDPTVTLNKGAVFVLT-----KQVWVSG-----PGNIAFNEICNGV 857
QY 492 KAVSNAAKOSFNPFAOEAPLVNPLMGSTIDVPRPPONFIEIGTE---CAPYEKRRVAVG 548
Db 858 AHAIAVDISIFE--NASIGASLF--LLSGPPLDVLITIKSTVGNSTVDNFNAPI---LVVSG 911
QY 549 ISAVTLRHSRGHGRKFRHVAVGASFRMPCGDTLSLGAOLFAARDKXYFMNTN----- 604
Db 912 IDSM-----INNGOYIGDQKNI-----IALSLG-----SDNSITVNSNTLYA 948
QY 605 ---FAATYAGSLRLQHDASLYSVSILLEGGLREILPPVSKTLPCSFYGOLSYGHTHD 661
Db 949 GIRTKTNOGTVLT-----SGGI-----PNNPQITYG-LGLENGDP 983
QY 662 RMKTESLPPPTPLSTDHSTWCGYVWAGELGTVAVEN-----TSGRGFRRETIPEV 713
Db 984 KIK-----QVTFETDYNMIGSII-----ATVVTINDVTLTGTGAGTDF----- 1023
QY 714 KVOAVYSRODSFVELAISRDFSDSHLYNLAIPLGIKLEKFAEOYHVAVMTSPDYCRS 773
Db 1024 -----DGKTTLSIGN-----ANVKFYDKRT-----S 1046
QY 774 NPKCTTLLS---NOSGSMKSGNLARQAGI---VOASGRFSLGAAMELFNGFEMWRGSS 828
Db 1047 HP--TSMIVSTKANOGVYVLYGNALVGNIGSSDIPVASVFTG-----NDSGVGLQGN 1098
QY 829 RSYNVDAQS 837
Db 1099 HSONIDFET 1107

RESULT 5
EGF2_YEAST
ID EGT2_YEAST STANDARD; PRT: 1041 AA.
AC P42835;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE EGT2 PROTEIN PRECURSOR (EARLY G1 TRANSCRIPT 2).
GN EGT2 OR YNL327W OR NO320.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / FY1676;
RX MEDLINE=95373280; PubMed=7645347;
RA Maftahi M., Nicaud J.-M., Levesque H., Gallardin C.;
RT "Sequencing analysis of a 15.4 kb fragment of yeast chromosome XIV
RT identifies the RPD3, PAS8 and KRE1 loci, five new open reading
RT frames".
RL yeast 11:567-572(1995).
RN [2]
RP SEQUENCE FROM N.A. AND FUNCTION.
RX MEDLINE=96251274; PubMed=8668141;
RA Kovacech B., Nasmyth K., Schuster T.;
RT "EGT2 gene transcription is induced predominantly by Swi5 in early
RT G1".
RL Mol. Cell. Biol. 16:3264-3274(1996).
CC -1- FUNCTION: SEEMS TO BE INVOLVED IN THE CORRECT TIMING OF CELL
CC SEPARATION AFTER CYTOKINESIS, AS SEPARATION OF MUTANT DAUGHTER
CC CELLS IS DELAYED. COULD EITHER BE AN ENZYME NECESSARY FOR GLICANS-
CC DEGRADATION OF THE CELL WALL AT THE NECK REGION BETWEEN MOTHER AND
CC DAUGHTER CELLS OR A REGULATORY PROTEIN CONTROLLING THIS METABOLIC
CC STEP. EXCLUSIVELY EXPRESSED BETWEEN THE END OF MITOSIS AND EARLY
CC G1; INACTIVATED BEFORE CELLS PASS START.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
DR EMBL: 246259; CAAB6371.1; -
DR EMBL: 271603; CAAG6259.1; -
DR SGD: S0005271; EGT2.
KW Glycoprotein; Repeat; signal; Cell cycle; Cell division; Mitosis.
FT SIGNAL 1 20
FT CHAIN 21 1041
FT DOMAIN 200 203
FT DOMAIN 381 384
FT DOMAIN 388 395
FT DOMAIN 490 493
FT DOMAIN 586 589
FT REPEAT 457 492
FT REPEAT 577 606
FT REPEAT 613 647
FT REPEAT 716 745
FT REPEAT 773 802
FT REPEAT 811 840
FT REPEAT 849 886
FT REPEAT 887 924
FT REPEAT 925 962
FT CARBOHYD 65 65
FT CARBOHYD 103 103
FT CARBOHYD 161 161
FT CARBOHYD 175 175
FT CARBOHYD 249 249
FT CARBOHYD 332 332
FT CARBOHYD 401 401
FT CARBOHYD 435 435
FT CARBOHYD 465 465
FT CARBOHYD 485 485
FT CARBOHYD 506 506
FT CARBOHYD 526 526
FT CARBOHYD 544 544
FT CARBOHYD 556 556
FT CARBOHYD 635 635
FT CARBOHYD 636 636
FT CARBOHYD 657 657
FT CARBOHYD 709 709
FT CARBOHYD 756 756
SQ SEQUENCE 1041 AA; 108494 MW; 01FECF8EAA8744CD CRC64;

Query Match 4.7%; Score 204; DB 1; Length 1041;
Best Local Similarity 21.5%; Pred. No. 0.00023;
Matches 168; Conservative 111; Mismatches 280; Indels 224; Gaps 37;

QY 10 ISLVPTLSMNLGAATTELSASNSFDTSTSTFS---SKTSATDGTNYVKDSVI 66
Db 374 MSAYTVSSSSBOASSSISLAPSSNSFTTPSSLSATETYSIISASISYTOASYI 433
QY 67 ENVPKGTGTOSTSCFKNDAAAGDLNPLGGSFTFSNIDATTAGS---AAGSEANKTV 123
Db 434 DNSGTTAVTOSTSTYI---AVSSAEKLSPTLSYR-SWVTISVSATGHTTTPSVSNST- 487
QY 124 TLGFSAL-SFLKSP--ASVTYNGLGALINVKGNLSLDNDKVLIONFSYGGGAINCAG 180
Db 488 TLSSSVTLSESVISSPYLANTVTSQASASQSTNPYPVSNSTSSATOLATAPAINITG 547
QY 181 ---SLKTAANKSLS-----FIGNSSSTRGAL---HTKMLTUSGGETL-----FQ 220
Db 548 TSISSSTTNSSVSSSTSSLSGCPVSNVAVASGYLLTTTBSAQLTEIGSLIPITIT 607
QY 221 GNTAPTAAGGGAIAIADSGTL--SIGDSGDILFEONTIGATGVG----- 265
Db 608 TSTTTSGTGTGSKNKVASSSTREIAQSIYVNSSLSVSTINTNAATPAANARNATFTHATSG 667

```

```

OY 266 -----HSAIDIGTS--AKITLRAAGHTIYEDPTVYGSTVADALNIN--SPDTGDN 316
DB 668 SIQPSVHSSSLSTIDTKYTTATFTSTR-----DGSSIAFTTGLNOSVYTGTD 717
OY 317 KEYGTIVTSGE--KLTEAFA-----KDEKNFTSKLLONAVFRNGTVLAKD 361
DB 718 KSPYEVSISSTESAQAQVEYDLSLPITLKPVTVTGTSRNFSTFMSVSTKLTEATATDKD 777
OY 362 VVLASNGFSQDANSKLIMDGLTSIVANTESEIETLNLEINIDLNRKKIKLSATAQKDI 421
DB 778 AY-----SVISSTQSAQVTEY-----GSMALIST-----801
OY 422 RIDRPVLAISDESEY-----QNGFLNDHSYDGLLELDAGKDIYISADRSINAVOS 474
DB 802 -LETPVIMSTDESGFTLTCTESGQATEYGLIPISLTDG--SVIYTFGESV-----853
OY 475 PYTGQKWTINMSTDDKKATVSAKOSFNPTAQDEAPLV--NLMSGFIDVRFQNFTEL 533
DB 854 -----VVGYSY-----TVGAQAQYA-----QHTSLVPSVSTIKS-----KTSL 885
OY 534 GTEGAEYERKFRVWAGISNVLHRSRGRENQKFRHVSAGAVGASTPMGCGDTLSLGFQOLF 593
DB 886 STEESV-----VAGSYTYGAQAQYAQHTSLVYVS--TIGSKTSLSTESVYAGVSTT--936
OY 594 ARDKYFEMNTN-----FAKTYA-----GSLRLQHDASLYSVVSIIL--GEG 632
DB 937 VDSQAQAEHTNLVAIDTLTKSTFOKATATEVCYTALSSPHSATILDAGTTISLPTSSST 996
OY 633 GIREILLPYVSKLPCSFYQQLSYGHTDHMKTESLPPPPPTSLTDHTSMGCVYVWAGELG 692
DB 997 SLSTITWTWSSSTI-----KPPSIST-----YSGA--AGQLT 1026
OY 693 TRV 695
DB 1027 IRI 1029

```

RESULT 6  
OMPR, RICRI STANDARD; PRT: 1654 AA.

AC 053047;  
DT 30-MAY-2000 (rel. 39, Last sequence update)  
DT 30-MAY-2000 (rel. 39, Last annotation update)  
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (POMP B)  
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
(120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].

GN OMPB.  
OS Rickettsia rickettsii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=783;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R;  
RC MEDLINE=92167802; PubMed=1724278;  
RA Gilmore R.D. Jr., Cleopla W. Jr., Pollicastro P.F., Hackstadt T.;  
RT "The 120 kDolalation outer membrane protein (Pomp B) of Rickettsia  
rickettsii is encoded by an unusually long open reading frame:  
evidence for protein processing from a large precursor.";  
RT Mol. Microbiol. 5:2361-2370(1991).  
RN [2]  
RP SEQUENCE OF 279-1654 FROM N.A.  
RC STRAIN=R;  
RC MEDLINE=90136087; PubMed=2515418;  
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;  
RT "Cloning, expression and sequence analysis of the gene encoding the  
RT 120 kD surface-exposed protein of Rickettsia rickettsii.";  
RT Mol. Microbiol. 3:1579-1586(1989).  
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL

```

CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X16353; CAA34403.1;
CC Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
CC CHAIN 1 1333
CC FT CHAIN 1334 1654 32 KDA BETA PEPTIDE.
CC DOMAIN 1181 1188 POLY-THR.
CC SEQUENCE 1654 AA; 168184 MW; D7AB70FB708F618 CRC64;

```

Query Match 4.7%; Score 203.5; DB 1; Length 1654;  
Best Local Similarity 22.5%; Pred. No. 0.00046;  
Matches 187; Conservative 102; Mismatches 349; Indels 193; Gaps 41;

```

OY 27 TEELASNSFDGTT-----TTFSSKTSSATDGTNVEFDSVYIENVPTGE-----74
DB 312 TANAGANAVIQTNNNGRAGAEVVSVDNGKVAITDG--OVYAKDVIQSANATGQVNR 369
OY 75 -----TQSTCFKNDAA-----AGDINFLGSGFSEFNSI---DATTAAGAAIGSEAN- 120
DB 370 HIYDVGADGTTAKRTASKYTTIODSNFNGTDFGNLAQIKVYRNALITLGNFNG--DASNP 428
OY 121 -KTVTLGFSALFLKSPAS---TYTNGGAINVKG---NLSDNDKVLIQDNFSTG 171
DB 429 GNTAGVITPDANGTLBSASADANVAATNNITALEAGAGVQVQSGTHAELRL-----G 482
OY 172 DGAINCAGSLKIANNKSLFTIGNSSSTRGGAHITNLTLSGCEFLFQ--GNTAPTAAG 229
DB 483 NAGSI-----FKLADGVIVKNGVQTAALVAGLAAGITL--DSASAITDIGNAGGAAL 536
OY 230 KGAIAIADSGTGISGDSDDIIFEGNTGATGTVSHALDLSAKITLRAAGCHTIV 289
DB 537 QRTTLNDAKKTLLTG-----ANIIGAG-----GTTDLANGCTIKLTSTONNIY 584
OY 290 FYDPIVTGTSYADALNINSPDGTGNKREYGTIVFSGEKLTEAKDEKNRTSKLLQNV 349
DB 585 DFLAATIDQTVGVVDSASLINAQTLFINGKIGTIGANNKTLGOFNIGSSKTVLSN--GNV 642
OY 350 AF-----KNGIVYLAGDVYLSANGFSQDANSKLI---MDLTSIVANTESEIETN--L 397
DB 643 AINELVIGNDGAQVAFADHYLLIRTTNAAQOGKTIIFNPVNNCTTLAAGTNLGSATNP 702
OY 398 EINIDSLNRKKIKLSAATFAOKDIRIDRPVLAISD--ESFYQNGFLNEDHSYDGLIELDA 456
DB 703 EINEFS-----KGVND--TVLVNVEGVNLTAINITTTDANV--GSVFVFN 744
OY 457 GKOIVISA-----DSRSINAVOSPYG---YQKWTINMSTDDKKATVSAKOSFNPTAE 507
DB 745 GGTIVISGVGGQGGKKNFTVVALENGTIVFLGNATFNCGNT--TIANSTLQIGCVYAD 802
OY 508 QEAP-----LYPMLMGSPFDVRFQNFTELCEGAPYE--KRFWVAGISNVL-----553
DB 803 CVASADGTGIV-----EPVNGPIT--VTLNKAQAPVNAKQITVSGPQNVVINEIGNA 854
OY 554 --HRSRGRENQKFRHVSAGAV-----GASTRMP-----GGDTLS---586
DB 855 GNHGGAVTTDIAFENSLGAVVFLPGIPNDAGNMPPLTIKSTGVGNKRAKGDVSVYV 914
OY 587 LGFQAOLFARDKDFEMNTNFAKTYAGSLR--LOHDASLYSVVSIILGEGGIRELLPYVSK 644
DB 915 LGVDVYIAQGVYIGDQNNITVGLGSDNGIIVATLVLAGISPLNNNGG--VTVLSGAVP 972
OY 645 TLPCSFYQQLSYGHTDHMKTESLPPPPPTSLTDHTSMGCVYVWAGEL-----GTRVAVENTS 701

```





```

FT CARBOHYD 330 330 N-LINKED (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (POTENTIAL).
FT CARBOHYD 436 436 N-LINKED (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (POTENTIAL).
FT CARBOHYD 799 799 N-LINKED (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (POTENTIAL).
FT CARBOHYD 826 826 N-LINKED (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (POTENTIAL).
FT CARBOHYD 861 861 N-LINKED (POTENTIAL).
FT CARBOHYD 879 879 N-LINKED (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (POTENTIAL).
FT CARBOHYD 1116 1116 N-LINKED (POTENTIAL).
FT CARBOHYD 1128 1128 N-LINKED (POTENTIAL).
FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
FT CARBOHYD 1146 1146 N-LINKED (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (POTENTIAL).
SQ SEQUENCE 1300 AA; 132801 MW; E09E52C3F647243D CRC64;

```

Query Match 4.5%; Score 193; DB 1; Length 1300;  
 Best Local Similarity 22.5%; Pred. No. 0.0013;  
 Matches 181; Conservative 91; Mismatches 317; Indels 216; Gaps 40;

```

QY 37 DGTSTSSSSKTSATDGT--NYVKDSVLENVKGTGTOSTSCFKDAA-----AGD 89
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
DB 23 DGTAAKTAAAKVTITQDSNFGNTDGNLAQIKVPA--ITLTGFTGDASMPGTAGV 80
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 90 LNFLGSGFSFTS-----NIDATPASAIAI-----GSEAAKVTLTSGFSALSFLKS 136
   : : : : : : : : : : : : : : : : : : : : : : : : :
DB 81 ITPDANGTLESASADANAVATNNITAIIESGACVOLSGTHAEFLGNG-----SIFKL 136
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 137 PASTVINGGAIIVKGNLSLNDKVLIDNFSTGDGAINCAGSLIKIANKSLS-FIGN 195
   ||| : : : : : : : : : : : : : : : : : : : : : : :
DB 137 ADGTVING-----KVNQNAL-----VGAL--AAGTITLDSATITGDIIGN 175
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 196 SSTRGCAIHTKNTLSSGGETLFOGNTAPTAAGKGAIALDSGLTSLSGSDGIIFES 255
   : : : : : : : : : : : : : : : : : : : : : : : : :
DB 176 A-----GGAALQRTILANDAKK-----TLTLGG-----A 200
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 256 NTIGATGVSHSIDLGTSAKITALRAAGHTIYFDPITVYSGTSVADALNINSPDTGD 315
   ||| : : : : : : : : : : : : : : : : : : : : : : :
DB 201 NIIGAGG-----GTIIDQANGGITKLSTONNIYVDDLAIDOTQVDDASSLTNQTLL 256
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 316 NKEYTGVIFSGEKLTEAEKDEKNTSKLQNVAF-----KNGTVVLKGDVLSANGF 369
   ||| : : : : : : : : : : : : : : : : : : : : : : :
DB 257 INKIGTIGANKKTLGOFNIGSSKTVLSN--GNVALINELVIGDGAIVQAHDIYLLTRTT 314
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 370 SODANSKL-----MDIGSLVANTESIELTN--LEINIDSLRNGKKIKLSAATAQDIRI 423
   : : : : : : : : : : : : : : : : : : : : : : : : :
DB 315 NAAQOKKIIFNPVNVNNGTTLAAGTNGSATNPPLAEINFGS-----KGVMV 359
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 424 DRPVVLAISD-ESEYNGNGLNEDHSYDGLLELDKADIVISA-----DSRSINAVQSPG 477
   ||| : : : : : : : : : : : : : : : : : : : : : : :
DB 360 D--TYLVANVEGNLTAFTNTTTTDANV--GSFVENAGGTNTIVSGVGQGNKFTVALENG 416
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 478 ----YOGKWTINMSSTDKKATVSWAKOSFNPTAEORAP-----LYPNLLMGSEFIDVRPQ 528
   : : : : : : : : : : : : : : : : : : : : : : : : :
DB 417 TTYKFLGNMTFNQNT--TIAANSTLQIGNYTAADCAVSDGTGIV-----EYVNTGPIIT 468
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 529 NTELEGTBGAPE--KRFVAGISNVL-----HRSGRNQRFRVHSGAVV-----573
   : : : : : : : : : : : : : : : : : : : : : : : : :
DB 469 --VTLNKAAPVVALKQITVSGPNVINEIGNAGNHGAIVDTITAFENSISGAVVFLRP 526
   : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 574 -----GASTRMP-----GGDTLS--LGFALFARDKDFMNTNFAKTYAGS 612
   ||| : : : : : : : : : : : : : : : : : : : : : : :
DB 527 GIPFNAGNTMPLTISYVGNKTAKEGDFVSVYLVGVDVSIACQVIGDQNNIVGLGLGS 586
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 613 LR--IQHDASLVSVSILLGEGGLREILLPYVSKTLPCEFSYGLSYGMDHRMKTESLPP 670
   : : : : : : : : : : : : : : : : : : : : : : : : :
DB 587 DNGIIVNATLVLVAGISTLANNQG--FVTLISGVPNPGFYVG--LGNGIGASKRK-----637
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 671 PPPTLSTDHTSMGQYVWAGEL---GTRVAVENTSGGFFRETPFPKVAQVYRSODSFVE 727
   : : : : : : : : : : : : : : : : : : : : : : : : :
DB 638 -QVFTTDDYNNLGNIIATNATINDGVTVTGTGIGIGF-----DCKIT 679
   : : : : : : : : : : : : : : : : : : : : : : : : :
QY 728 LGAISRD---FSDSHLYNLAIPLG 748
   ||| : : : : : : : : : : : : : : : : : : : : : : :
DB 680 LGSVNGNANRFADGILSNSTSMIG 704
   : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 10
YFJA.ECOLI STANDARD: PRT: 1569 AA.
AC P52143; P76610; P77017; P77019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 98.4 KDA PROTEIN IN ALPA-GABD INTERGENIC REGION (F949).
GN YFJA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Glasner J.D., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono K., Itoh T., Kanai K.,
RA Kasai H., Kashimoto K., Kim S., Kimura S., Kitagawa M.,
RA Kitakawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H.,
RA Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C.,
RA Yamamoto Y., Yano M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).
CC -I- SIMILARITY: STRONG, TO BORDETELLA PERTACTIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC
DB EMBL: U36840; AAA79815.1; ALT-SEQ.
DB EMBL: AE000350; AAC75695.1; -.
DB EMBL: D90889; BAA16514.1; ALT_INIT.
DB EMBL: D90890; BAA16518.1; ALT_INIT.
DB Ecogene: EG13213; YFJA.
KW Hypothetical protein; Outer membrane.
SQ SEQUENCE 1569 AA; 162774 MW; 0D006A9A32154596 CRC64;

```

Query Match 4.4%; Score 188.5; DB 1; Length 1569;  
 Best Local Similarity 22.0%; Pred. No. 0.0031;



Matches 124; Conservative 75; Mismatches 229; Indels 135; Gaps 26;

QY 19 SNLGAATTEELASNSFP--GTTSTSTSSKSSSATDGTNYFKSVIENPKTGENO 76  
 Db 76 TNYGVVNDIYDGNOKVBERGTTNTHIINHOQNVHGG--VSNGLI-----ESGGTQ 128  
 QY 77 STSCFKNDAAGDINLGGCFSTFSNIDATVATGAAGISSEANKTVTLISGFSALFLKS 136  
 Db 129 DIGSHNFPVQGANNTINGRO---SIHOGISTGTIT--ESGNQVYVGGISNGTTIG 183  
 QY 137 PASTVINGGALVKNLNLIDNDKVLIDQNSTGCGAINCAGSLKIANKSLSTIGNS 196  
 Db 184 GASRYVGG-----SANGILIDGSGQIVVQV--GHADTTINKSGSODVVOGS---LATN 232  
 QY 197 SSVRGGAH-----TKNLTLSSGCE-----TLFQG-----TFAPTAAG-----KGG 232  
 Db 233 TTINGRQVVEOSTVETTTIKNGEQRVYESRALDTTIEGQIOLNSKSTAKNTHIYSGG 292  
 QY 233 -----AIAIADSGTSLISGDSG-----DIEFGNTIGAT--GVYSHALIDLG 272  
 Db 293 TQIVDNTSTDVIEVYSGGLVDRGATATNVQHDGAILKTNTNGTTVSGTSEGAFSIH 352  
 QY 273 TSKKITALAAQGH--TITYYDPTTV-----GSTV-----ADALNI-----NSP 311  
 Db 353 NHVADNVLLENGHIDIMVYGSANKTIIKDKGTMSVLTNAKADATRIDNGGVADVAGNAT 412  
 QY 312 DTGDN-----KEY---TGTIVFSGEKTFEAKDEKNRTSKLLQWVAFKNGTVV----- 357  
 Db 413 NTINGCTONINNYGIATGTNTINSGONIKSGKADPTTIISGSRQVYERKDAISNITS 472  
 QY 358 LKGDVVLANSFGSDANSKLIINDLGTSLVANT-----ESIEPTNLEINISLR 405  
 Db 473 AGGSLIVYGGIAHGVN-----QETGSALVANTGAGTDIEGYNKLSHFTITGGEANVYLE 528  
 QY 406 NCKKILSAATFQAKOIRIDRPVVLASDSFYONGFLNEDHSDGLLELDAGD----- 459  
 Db 529 NGELIVLVAKTSKNTTIDTGKGLIYQKRAKTDSTRLN---GGLVEVQDGGEAARHVEQ 584  
 QY 460 -----IVISADSRISINAVOSPYG 477  
 Db 585 QSGGALIASTTSGTLLEGTSNYG 607

RESULT 11  
 ID 190K\_RICRI STANDARD; PRT: 2249 AA.  
 AC P15921;  
 DT 01-APR-1990 (rel. 14, Created)  
 DT 01-APR-1990 (rel. 14, Last sequence update)  
 DT 01-OCT-1994 (rel. 30, Last annotation update)  
 DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).  
 OS Rickettsia rickettsii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiense; Rickettsia.  
 NCBI\_Taxid=783;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=R;  
 RX MEDLINE=90354033; PubMed=2117568;  
 RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.,  
 RT "A protective protein antigen of Rickettsia rickettsii has tandemly  
 RT repeated, near-identical sequences."  
 RL Infect. Immun. 58:2760-2769(1990).  
 CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY.  
 CC -1- PFM: GLYCOSYLATED (POSSIBLE).  
 CC -1- DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF  
 CC AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M31227; AAA26380.1; -  
 DR PIR: A41477; A41477.  
 KW Antigen; Repeat; Signal; S-layer; Glycoprotein.  
 FT SIGNAL 1 28  
 FT CHAIN 29 2249  
 FT DOMAIN 212 1180  
 FT REPEAT 212 286  
 FT REPEAT 287 358  
 FT REPEAT 359 430  
 FT REPEAT 431 505  
 FT REPEAT 506 577  
 FT REPEAT 578 652  
 FT REPEAT 653 724  
 FT REPEAT 725 799  
 FT REPEAT 800 874  
 FT REPEAT 875 949  
 FT REPEAT 950 1021  
 FT REPEAT 1022 1093  
 FT REPEAT 1094 1165  
 FT REPEAT 1166 1180  
 SQ SEQUENCE 2249 AA; 224333 MW; A9D6646C089DF087 CRC64;

Query Match 4.4%; Score 187; DB 1; Length 2249;  
 Best Local Similarity 24.1%; Pred. No. 0.0063;  
 Matches 154; Conservative 89; Mismatches 282; Indels 114; Gaps 31;

QY 15 TISMNLGAATTEELASNS--FDGTTSTSTSSKSSATDGTNYFKSVIENPK 71  
 Db 302 FATLGAIVAKATTTKLTLDAASAKFTNPVVTGAIQNTGNANNGIYFTGNSVTGCVN-- 359  
 QY 72 TGTOSTSCFKNDAA-----AGDLNPLGGGCFSTFSN-----IDAFTTSGAII 114  
 Db 360 -GNTNLATVNVGAGLLOVQGVYKANTNLTDNASAVFTNPVVVTGADTGNANNGI 418  
 QY 115 GSEPAKTVT--LSGFSALFLKSPASVTVNGIGAI-----NVKGNLSLNDKVL 163  
 Db 419 VFTFGNSTYTGDIQNTNATATVNVGAGTILG--GAVIKATTTKLTNAASVLTLTNANAVL 477  
 QY 164 IQ--DNFSTGDC--GAINCAGSL-----KIANKSLST--GNSSSTRGCAIHTKNTLSS 213  
 Db 478 TGAIDNTTGGDNVGVNLGALSOVYTGNTGNISLATISVGAGTATLGGAVIKATTTKLT 537  
 QY 214 GGETLFGQNTAPPAAGKGAIAIADSGTSLISGDS--GDIIFEGNT-----IGATGT 263  
 Db 538 DAASAVKFTNPVVTGAIQNTGNANNGIYFTGNSVTGDI--GNTNSLATISVGAGTA 594  
 QY 264 VSHSAIDLGTSAKITALRAAQCHTIFYDPI-----TYTGSTVADALNIN--SPDTG 314  
 Db 595 TLGGAIVIKATTTKLT--NAASVLTLTNANAVLTGAIQNTTGGDNV--VNLNGALSOVYTG 651  
 QY 315 D-----NKEYTGTVSGEKLTEAEAKDEKNRTSKLLQNAFAKNGIYVLKGDVVLANSNGS 370  
 Db 652 DIGNTNSLATISVGAGTATLGGAVIKATTTKLTNAASAVKFTN--PVVTGAIQNTGN-- 707  
 QY 371 QDANSKLIDLGTSLV-----ANTSEITELNLEINISLNGKKIKSATFQAKOIRIDRP 426  
 Db 708 --ANNQIVFTGNSVYTGDIQNTNATATVNVGAGTILG--GAVIK--ATTK--LTNA 759  
 QY 427 VVLASDESFYONGFLNEDHSD--GILELDAGKDIVISADSRISINAVOSPYGOGKWTI 484  
 Db 760 SVLTLTNANAVLTGAIQNTTGGDNVGLNLN--GALSQVYTGDIQNTNSLATISVGAGTATL 818  
 QY 485 NMSTDDKATVSMAKOSFNPABQAPLVNPLMKSEFIDVRPQNFIEGTGAPPEKRF 544  
 Db 819 GGAV--IKATTTKLTNAASVLTLTNANAV--LTGAVDNTTGGDNVGLNLNGA----- 867  
 QY 545 WVAIGSNVLRBSRGQRKFRHVS-----GGAVYGAST 577  
 Db 868 ---LSQVYTGDIQNTNSLATISVGAGTATLGGAVIKATTT 902

RESULT 12  
 OMPB\_RICPR STANDARD; PRT; 1643 AA.  
 AC 053020; 09ZCMT0;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
 DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).  
 GN OMPB OR SPAV OR SPA OR RP704.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiinae; Rickettsia.  
 NCBI\_TaxID=782;  
 [1]  
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-BREINL;  
 RA MEDLINE=91045972; PubMed=2122457;  
 RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;  
 RT "Characterization of the gene encoding the protective paracrystalline-  
 RT surface-layer protein of Rickettsia prowazekii: presence of a  
 RT truncated identical homolog in Rickettsia typhi";  
 RT Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-BREINL;  
 RA Moron C.G., Yu X.J., Walker D.H.;  
 RT "Sequence analysis of ompB of Rickettsia prowazekii";  
 RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RA MEDLINE=99039499; PubMed=9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Scharnberg-Ponten T., Alsmark U.C.W., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria";  
 RT Nature 396:133-140(1998).  
 RL [4]  
 RN PARTIAL SEQUENCE.  
 RC STRAIN-BREINL;  
 RA MEDLINE=92114896; PubMed=1370573;  
 RA Ching W.M., Carl M., Dasch G.A.;  
 RT "Mapping of monoclonal antibody binding sites on CNBR fragments of  
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
 RT prowazekii";  
 RT Mol. Immunol. 29:95-105(1992).  
 RL [5]  
 RN IDENTIFICATION OF CLEAVAGE SITE.  
 RA MEDLINE=92104668; PubMed=1729180;  
 RA Hackstadt T., Messer R., Cleplak W. Jr., Peacock M.G.;  
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer  
 RT membrane protein of rickettsiae: identification of an avirulent mutant  
 RT deficient in processing";  
 RT Infect. Immun. 60:159-165(1992).  
 RL [6]  
 RN FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC [7]  
 CC FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
 CC [8]  
 CC SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY.  
 CC [9]  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC  
 DR EMBL: M37647; AAA26390.1; ALT\_INIT.  
 DR EMBL: AF161079; AAD42234.1; -  
 DR EMBL: AJ235273; CA15140.1; -  
 KW Antigen; S-layer; Cell wall.  
 FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.  
 FT VARIANT 257 257 V -> A (IN STRAIN BREINL).  
 FT VARIANT 1010 1010 Y -> A (IN STRAIN BREINL).  
 FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).  
 FT VARIANT 178 179 AA -> VC (IN REF. 1).  
 FT CONFLICT 191 201 TFOEAPLTGA -> INSRSSYHLVS (IN REF. 1).  
 FT CONFLICT 212 212 T -> I (IN REF. 1).  
 FT CONFLICT 313 313 Q -> L (IN REF. 1).  
 FT CONFLICT 1104 1104 D -> G (IN REF. 2).  
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).  
 SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;  
 Query Match 4.38; Score 186; DB 1; Length 1643;  
 Best Local Similarity 21.28; Pred. No. 0.0047; Mismatches 302; Gaps 52;  
 Matches 199; Conservative 127;  
 16 LSNMNLGAATTEELASNSFGDTSTTSFSSKT---SATQCTNVFVKDSVIVNPK 71  
 DB 356 IOSANAGGVTFEHI-VDVGLGTTNFKTADSKYITENSFGSTNGNDIQIV--VPD 412  
 QY 72 T-----GETQSTCFKNDAAAGDNLFLGGF-----SFTFSNIDATTAAGAIG 115  
 DB 413 TKILKNGFIDVKN-----NGNTAGVITFNANGALVASTDPNIAVININAEAGAGV- 466  
 QY 116 SEAAKTVTLISGF-----SALSFLKSPASTYTNGLGAIN--VGNLSL-----LD 158  
 DB 467 -----VELSGIHIAELRLNGSGSIFKLDAGVYN--GPNONALMNNNALAGSITOLD 517  
 QY 159 NDKVLIQDNSTDPDGCALNCAGSLKIAN--NKSLSFGNS--SSTRGAGIHTKMLTLSSG 214  
 DB 518 GSAIITD---IGNGVNAALQHTTLANDASKILALDGANITIGANGVAIH-----FOAN 569  
 QY 215 GETLFGQNT-----APTAAGGGAIAIADSGTSLISGSDGIIIEGNT----- 257  
 DB 570 GGTIKLTNTQNNIVNFDLITTDKGTGVDASSLTNNQTLTNGISGTVAANKTIAQLN 629  
 QY 258 IGATGY-----SHSAILDGLSAKI---TALRAAGHTIYYDPPIYVIGST 300  
 DB 630 IGSSKTLNAGDAVAINELVETNNNGSVQLNHTYLTITKINAANOGQIIVAADPLNT--NT 687  
 QY 301 SVADALINSPDGDNKEKYGTTFVSGEKLTEAPAKDEKNTSKLQNVAFKNGTVVLKG 360  
 DB 688 TLADGTLGSAENP-----LSTHFA---TKAANDS-----ILNVC-----KG 723  
 QY 361 DVLVLANGFSQDAN--SKLMDLQTSLVANT---ESIELTNEINIDSLRNGKKIR-LS 413  
 DB 724 VNIYANNITINDANVGLSHFRSGTSIVSGVGQOQHKLNNL-----IIDNGTVYFVLG 778  
 QY 414 AATRAQDIRIDRPVVALISD-----ESFYQNG-----FLNDDHSDGLLEL 454  
 DB 779 DTFENGSTKIEGKSIILQISNNYTTDHVESADNQTLEFVNTPPIYVILNAGAFGLV- 836  
 QY 455 DAGKDIYISADS-----RSINAYOSPYGQKWTINWSTDKKATVSAKOSFNPTABQEA 510  
 DB 837 ---KQYIISPGNIVFNEIGNVGIVHGIAN-----SISENNSLGLSL----- 877  
 QY 511 PLVFNILMGSFIDVRPQNFIELTE--GAPYEKRPWAGISNVLHRSRGRENORFRHV 567  
 DB 878 -FLPS---GAPLDVLTFTKSVNGTVDNFMNAPI---VVVSGIDSM-----I 916  
 QY 568 SGAAVVGASTRMAGGGLSLGFPQLPARDDYFMNTFATYAGSLRLQHDASISYVSI 627  
 DB 917 NNQIILIDKKNI---IALSLG-----SDNSITVYNAV-----TLYSGIRT 952  
 QY 628 LLEGGLREILPLVSVSKLPCSFYGLSYGHTDRMKTESLPPPPPTLSTDHNSMGGYW 687

```

Db 953 TKNNOS--TVTLGGMPNNFGTLYG-LGLENGSPKLR-----QYFTTDYNN----- 996
QY 688 AGEIAGRVAVENTSGKGFREYTPFKVQAVYSGRODSFEIAGISRD---FSDSHLYNL 743
Db 997 ---LGSIIANNVT-----INDIYTLTTGGIAGTDFDKITLGSVGNANRVYDS----- 1043
QY 744 APLPLIKERFPAEQYHYVAVYSPDVCRSNPACTT---TLISNOGSMKTKGSNLARQA 799
Db 1044 -----TFSDPRSMIYA-----TQANKGTVTYLGALVSNIGSLDTP----- 1079
QY 800 GIYQASGFRSLGAALFNGFNFGEMRGSSRSYVADGS 837
Db 1080 --VASVFTGNDGAGILOGNI-----YSQNIIDFGT 1107

RESULT 13
OMP, RICJA STANDARD; PRT; 1656 AA.
ID OMP, RICJA
AC 006653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_Taxid=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein romp B of Rickettsia
RT japonica."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -I- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AB003681; BAA20138.1;
CC Antigen; S-layer; Cell wall.
CC CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.
CC FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.
CC FT DOMAIN 528 533 POLY-GLY.
CC FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;
SQ

```

Query Match 4.3%; Score 184; DB 1; Length 1656;  
 Best Local Similarity 21.3%; Pred. No. 0.0061;  
 Matches 169; Conservative 92; Mismatches 254; Indels 278; Gaps 42;

```

QY 38 GTTSTTSSFKTS--SATDGTNY--VFKDSVYIENVPTGETOSTSCFRKND-----AA 87
Db 375 GIDGTATFATASIVAITGNSNGTDFGNLAQVYVPPT--MTLGNFTGDANNNGNNA 432
QY 88 GDLNLTGGGFSFTFS-----NIDATTASGAIGSEANAKTVTLGSPSALSLFKSPA 138

```

```

Db 433 GVITTFANGTLASASADANVATNNNTTAEASGVAVQLSGTTHAEIRLGNMAGSVERLAD 492
QY 139 STVING-----LGAIVNGKSLDNDKVLIDQDFSTGCGAICAGSLKANKKSLS 191
Db 493 GTVINGKVNQTVLVGGVLAAGAITL--DSATITTGIDIGNGGGA--ALQSITLAND----- 544
QY 192 FIGNSSSTRGCAIHTNRLNLLSSGGETLFGCN-----TAPTAAGK 231
Db 545 --ATKTLTLTGGCA-----NITSANGCTINFQANGCTIKLTSTONNIIVVDCDLATIDQTGV 598
QY 232 GATAIADSGTSLISGDSGLIFEGNT-----IGAT-----GTVSHSIDGTSKRT-- 277
Db 599 DASSLTNAQTLTITSGIG--IIGANNITLGOFPNIGSSKTLTNGNVAINELVIGNNNSVGF 657
QY 278 -----TALAAGQHTIYFDPIYVGTSTVADALNINSP-----DTGDNKEYG 321
Db 658 AHNTYTLITTTNAGGCKII--FNPV--VNNNTTLAAGTNIGSAANPLAEINGSKARAD 714
QY 322 TIVFSGE-----KLTEAEAK-----DEKNRTS-----KLQNVAFKNGTVV- 357
Db 715 TVLVNVEGVALYATNTTTDDANVGSFVNAGKNIVSGVGGQGNKFNVTALDNGTVYK 774
QY 358 -----LKGDVYSANG--FSODANSKILMDL-GISLV--ANTESIELTNEINISLSR 405
Db 775 FLGNATFNGMTTIIAANSTLQISGNTYADFLASADGIVGEVNTGINTV-----L 825
QY 406 NGKKIKLSAATAQKDIRIDP--VVLAIODESFYONGFLN-----BDHSYDGLLEL----- 454
Db 826 NKQAVPVNAL--KQITVSGPCNVVYNEIGNAGNIVGAMTDTIATFNSSISLAVLFLPSGI 882
QY 455 --DAGKDIYISADSRISINAVOSPYGQKMTINWSTDKKATVSAKOSFNPAQDEAP 511
Db 883 PFNDAGNTIPLTI-----KSTVVG-----NETAGFS- 908
QY 512 LVNLLMGSTIDVRPFQNFLELTGEGAPYEKRWVAGISVILHRSRENDRKRHHVSGCA 571
Db 909 -VPSVI-----VSGVDSV-----IADQ 925
QY 572 VVGSATRPMPGDDTLISGFQALFARDQYFMNTVFATYAGSLRLQDASLVSYSILLGE 631
Db 926 VIGDQNNIYG--LGIG-----SDNGIYNATLIVAGIGINNQGVTL-----S 968
QY 632 GGRREILLPVSKTLPDSCFYGLSYGHTDHMKTESLPPPPPTLSLSDHTSWGCVYAGEL 691
Db 969 GGV-----PNTPGTYVG-LGTGIGASKFK-----QYFTTDYNN-----L 1002
QY 692 GTRVAVENTSGRG 704
Db 1003 GNIIATVNTINDG 1015

```

```

RESULT 14
ID AG43_ECOLI STANDARD; PRT; 1039 AA.
AC P39180; P76360; P75614; P97241; Q46771;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ANTIGEN 43 PRECURSOR (AG43) (FLUFRING PROTEIN).
GN FLU.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;

```

RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-97251358; PubMed-9097040;  
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
 Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,  
 Mizobuchi K., Mori T., Motomura K., Nakade S., Nakamura Y.,  
 Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,  
 Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,  
 Yamamoto Y., Horiuchi T.;  
 RA "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 40.1-50.0 min region on the linkage map.";  
 RL DNA Res. 3:379-392(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ML 308-225;  
 RX Henderson I.R., Owen P.;  
 RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP PRELIMINARY SEQUENCE OF 53-78.  
 RC STRAIN-ML 308-225;  
 RX MEDLINE-89291704; PubMed-2661530;  
 RA Caffrey P., Owen P.;  
 RT "Purification and N-terminal sequence of the alpha subunit of antigen  
 RT 43, a unique protein complex associated with the outer membrane of  
 RT Escherichia coli.";  
 RL J. Bacteriol. 171:3634-3640(1989).  
 RN [5]  
 RP SEQUENCE OF 53-63.  
 RC STRAIN-K12 / EMG2;  
 RX MEDLINE-97443975; PubMed-9298646;  
 RA Link A.J., Robinson K., Church G.M.;  
 RT "Comparing the predicted and observed properties of proteins encoded  
 RT in the genome of Escherichia coli K-12.";  
 RL Electrophoresis 18:1259-1313(1997).  
 RN [6]  
 RP GENE NAME.  
 RX MEDLINE-97257509; PubMed-9103983;  
 RA Henderson I.R., Meenan M., Owen P.;  
 RT "Antigen 43, a phase-variable bipartite outer membrane protein,  
 RT determines colony morphology and autoaggregation in Escherichia coli  
 RT K-12.";  
 RL FEMS Microbiol. Lett. 149:115-120(1997).  
 RN -1- FUNCTION: CONTROLS COLONY FORM VARIATION AND AUTOAGGREGATION. MAY  
 CC FUNCTION AS AN ADHESIN.  
 CC -1- SUBUNIT: HETERODIMER OF THE PERIPHERAL MEMBRANE PROTEIN (ALPHA  
 CC CHAIN) ANCHORED TO THE INTEGRAL OUTER MEMBRANE PROTEIN (BETA  
 CC CHAIN).  
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE ASSOCIATED.  
 CC -1- SIMILARITY: TO ADHESIN AID-I AND TO BORETELLA PERTACTIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 CC EMBL: AE000291; AAC75061.1; ALT INIT.  
 DR EMBL: D90838; BAA15825.1; ALT INIT.  
 DR EMBL: D90839; BAA15832.1; ALT INIT.  
 DR EMBL: U24429; AAB47869.1; -.  
 DR EcoGene: E612686; flu.  
 KW Outer membrane; Signal.  
 FT SIGNAL 1 52  
 FT CHAIN 53 551 ANTIGEN 43 ALPHA CHAIN.  
 FT CHAIN 552 1039 ANTIGEN 43 BETA CHAIN.  
 FT VARIANT 2 2 K -> N (IN STRAIN ML 308-225).  
 FT VARIANT 41 42 SL -> FF (IN STRAIN ML 308-225).  
 FT VARIANT 46 46 T -> K (IN STRAIN ML 308-225).  
 FT

FT VARIANT 157 157 W -> L (IN STRAIN ML 308-225).  
 FT VARIANT 188 188 V -> F (IN STRAIN ML 308-225).  
 FT VARIANT 303 305 ATN -> STI (IN STRAIN ML 308-225).  
 FT VARIANT 320 320 A -> T (IN STRAIN ML 308-225).  
 FT VARIANT 372 372 N -> Q (IN STRAIN ML 308-225).  
 FT VARIANT 493 493 E -> V (IN STRAIN ML 308-225).  
 FT VARIANT 497 497 S -> N (IN STRAIN ML 308-225).  
 FT VARIANT 585 585 H -> Y (IN STRAIN ML 308-225).  
 FT VARIANT 709 709 E -> K (IN STRAIN ML 308-225).  
 FT VARIANT 721 721 M -> T (IN STRAIN ML 308-225).  
 FT VARIANT 751 753 GHL -> SHE (IN STRAIN ML 308-225).  
 FT VARIANT 803 803 S -> P (IN STRAIN ML 308-225).  
 FT VARIANT 815 815 A -> V (IN STRAIN ML 308-225).  
 FT VARIANT 824 824 C -> S (IN STRAIN ML 308-225).  
 FT VARIANT 829 835 LNVHTS -> MNLVHTS (IN STRAIN ML 308-225).  
 FT VARIANT 845 847 QGT -> LCA (IN STRAIN ML 308-225).  
 FT VARIANT 855 855 S -> T (IN STRAIN ML 308-225).  
 FT VARIANT 888 888 Q -> L (IN STRAIN ML 308-225).  
 FT VARIANT 1025 1025 S -> I (IN STRAIN ML 308-225).  
 FT CONFLICT 61 63 ETY -> TTT (IN REF. 5).  
 SQ SEQUENCE 1039 AA: 106841 MW: 517064708DEBEO CRC64;  
 Query Match 4.2%; Score 182; DB 1; Length 1039;  
 Best local Similarity 21.6%; Pred. No. 0.0042;  
 Matches 162; Conservative 99; Mismatches 236; Indels 252; Gaps 43;  
 QY 23 GAATTEELASNSFDTSTT-----SFSSKTSATDGTNYEKSVYIENVPR---71  
 DB 301 GPTATNTLKGAGALVSTATAVTGINRLGAFSVEGKA-----DNVYLENGRLDV 351  
 QY 72 -TGETSTSCFRKDAAGDINFLGGEFSPFS--NIDATTA-SCAAI-GSEANKRYTILS 126  
 DB 352 LTHST-ATNTRVDD--GGTLDVRNGGTATVSMNGGVLLADSGAAVSGRSGKAPISIG 408  
 QY 127 GFSAFLSKSPASTVINGLG---AIVKGNLSLNDKVLIDNFSTGCGAINCAGSL 182  
 DB 409 GGGADALMLEKSGSFTLMGDTATDTVNGL-----FTANGGL--AGTT 452  
 QY 183 KIANNKSLSEFTGNSSSTRGCAIHTKNTLTSSGGETLFGQWTPAAGKGAIAIADSGTL 242  
 DB 453 TLNNGAILTLISGKT-----VNNDLTIRE-GDALOGGSLP-----GNGSVEKSGSGTL 500  
 QY 243 SIIG-----DSGDITFEGNTIGATGTVSHSADLTGSAKITALRAOQHITTFYD 292  
 DB 501 TVSNITTLTKRAVNLNESTLPLNDSTV--TTDVIAR-----GTALKLGSTVLNG---AID 551  
 QY 293 PIVTGSTVADALININSPDTG-----DNKEYTGTIVFSGEKLTEAEKDKNRKSTKL- 345  
 DB 552 PTNVT-----LASGATNRIIPNATVQSVVDLSHAGQIHFT-----STRCKFV 595  
 QY 346 -----LQNAFKNGTVLVKGVVLSANGFSODANSKILMDLGTSLVANTESIELTLEIN 400  
 DB 596 PATLTKNKLNGQGTISLR-----VRPDMAQNNADRVIDG-----RATGKTLNL-VN 644  
 QY 401 IDSLRNGKTKKLSAATRAOKIRIDRPVYLAISDESPQNFQNFLEMDHSYDILBLDAGKDI 460  
 DB 645 AGNSAGS-----LATVSGKIQ-----VVEAI-----NATTEEGFVQGNRLQAG--- 664  
 QY 461 VISADRSINAVQSPYQYQKWTINNSTDDKKAIVSAKQSFPTAEQEAFLVNLNMG 520  
 DB 685 -----AFN-----YSLNDSDE-----SWYLRSEN-AYRAEPLVLAASHL- 717  
 QY 521 FIDVRPFQNEIELGTEGAPYKRFVWAGISNVLHRSQENQKRFHVS-----G 569  
 DB 718 -----TQAMDYDR--IVAQSRS--HQGVNENNSVRLSIGGLGHQDNG 759  
 QY 570 GAVVGASTRMPGGDTLISGAQOLFARDKQVFMTNTRAKTITAGSLRLQHDASLSVYSIIL 629  
 DB 760 GIARGATPESG--SYGVRV----- 778  
 QY 630 GEGGLREIILPYVSKTLPCSFYQOLSYGHTDHRMKTESLPPPPPTLSTDHTSWGXY--- 685

DB 779 -EGDMLRTEVAGMSVT--AGVYG--AAGHSVYVKDD--GSRAGTVRDDAGCGLYLNTLV 832  
 OY 686 -----VMAGEL--GTRVAVENTSGRGPFR 707  
 DB 833 HTSSGIMADIVAOGTRHSMKASSDNNDFR 861

RESULT 15  
 SLAP CAUCR  
 ID SLAP CAUCR STANDARD: PRT: 1025 AA.  
 AC P35828; Q46015;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 30-OCT-1994 (Rel. 30, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE S-LAYER PROTEIN (PARACRYSTALLINE SURFACE LAYER PROTEIN).  
 GN R5AA.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter.  
 OX NCBI\_TaxId=76;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=93007489; PubMed=1393820;  
 RA Gilchrist A., Fisher J.A., Smit J.K.;  
 RT "Nucleotide sequence analysis of the gene encoding the Caulobacter  
 RL crescentus paracrystalline surface layer protein.";  
 RL Can. J. Microbiol. 38:193-202(1992).  
 RN [2]  
 RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=89008089; PubMed=3049545;  
 RA Fisher J.A., Smit J.K., Agabian N.;  
 RT "Transcriptional analysis of the major surface array gene of  
 RL Caulobacter crescentus.";  
 RL J. Bacteriol. 170:4706-4713(1988).  
 RN [3]  
 RP CHARACTERIZATION.  
 RC STRAIN=CB15;  
 RX MEDLINE=98292737; PubMed=9620954;  
 RA Avram P., Smit J.K.;  
 RT "The Caulobacter crescentus paracrystalline S-layer protein is  
 RL secreted by an ABC transporter (type I) secretion apparatus.";  
 RL J. Bacteriol. 180:3062-3069(1998).  
 CC -I- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY  
 CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.  
 CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A  
 CC S-LAYER WITH HEXAGONAL SYMMETRY.  
 CC -I- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE  
 CC SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A  
 CC SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF062345; AAC38665.1; -  
 DR HSRP: P01549; 2MCM.  
 DR InterPro: IPR001343; -  
 DR Pfam: PF00353; hemolysinCbind; 1.  
 DR PRINTS: PR00313; CABDNCRPT.  
 KW Cell wall; S-layer; Calcium-binding.  
 FT INIT\_MET 0  
 SO SEQUENCE 1025 AA; 98209 MW; AFC8B519820B1A5F CRC64;

Query Match 4.2%; Score 178.5; DB 1; Length 1025;  
 Best Local Similarity 22.5%; Pred. No. 0.0066;

Matches 121; Conservative 58; Mismatches 197; Indels 161; Gaps 22;  
 OY 10 ISLVPTLSMNLGAATTEELASNSFDDGTSTTSPSSKTS-----SA 52  
 DB 312 ISGIETMNTS--GAAT-----LNTSSGVTGTLANTNTSGAAGCTVTAAGONLATA 364  
 OY 53 TDGTNYVFKDSVVIENVPKTGETOSTSCFRNDAAADLFLGGSEFTTSDATYASGA 112  
 DB 365 AQANNVAVDGRANVTVAAGTGTGTTVGANSAA-----SGTVSVVANSSTTTTGA 418  
 OY 113 AI-----GSEANKVTVLGSFALSFLKSPASTVNTGLA---INVKGNLSLD 158  
 DB 419 AVTGTAVTVVAGTANAVNTTLQADVTVTGNSTTAVTVTQTAATAGATYAGRV---- 474  
 OY 159 NDKVLIODN-----FSTGDGALINCAGSLKIANKSLFIGNSSSTRGAI- 204  
 DB 475 NCAVTTTDSAAASATAGKATATVTLGSPGATITDSSALTYVN--LSGTGSLGIGGALT 532  
 OY 205 ---HTKNLFLSSGGETLFDGNTAPTAKG-----AIALDSGLTSLIS 245  
 DB 533 APPTANTLTLNVLNGLTFTTGATIDSEAAADGFTTINAGSTASSIASHVAADATLNTIS 592  
 OY 246 GDSGDIIFE-----GNTIGA-----TGYSHSAIDGTSKAITALA 282  
 DB 593 GDARVTITSHTAALTGITVTVNSVGATIDAEALATGLVFTGGAGRDSILGATTKAIVMGA 652  
 OY 283 AQGHTIYFDPITVGTSTVA-----DAL--NINSPDGDNKEYGTIVFSGEKL 330  
 DB 653 GD-----DVTYVSATITAGAGSVNGGDTDLVANVNGSSSADPARGG--FETLAV 702  
 OY 331 TEAEANDEKNRT-----SKLONVAFKNGTIVVAKGD-----VVLASNG---- 368  
 DB 703 AGAAAGSHNANGFTALQCATAGATTFTNVAVNGLVTLAAPTGTVTYVLANATGTSDV 762  
 OY 369 FQDANSKILMDIGTSLVNTESIEL-----TNLEINIDSLRNGKRIKLSAATAQXDI 421  
 DB 763 FLLTSSAALAGTVALGVEVTNIAATDTTAAHVDL-----TLQATSAKSI 812

Search completed: October 2, 2001, 03:32:28  
 Job time: 263 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2001, 03:31:13 ; Search time 119.05 Seconds  
(without alignments)  
934.637 Million cell updates/sec

Title: US-09-446-677B-8  
Perfect score: 4298  
Sequence: 1 MKIFLRFLLISLVPTLSMSN.....FEMRGSSRSYNDAGSKIKF 841

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%  
Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP unclassified:\*  
13: SP vertebrate:\*  
14: SP virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4278	99.5	841	2	0923A1
2	3394	79.0	712	2	09RB73
3	1495	34.8	839	2	P77792
4	1482	34.5	847	2	P71132
5	1429.5	33.3	928	2	086164
6	1429.5	33.3	949	2	09K299
7	1413.5	32.9	930	2	092393
8	1413.5	32.9	930	2	09RB65
9	1400.5	32.6	928	2	09RB65
10	1381.5	32.1	846	2	P71133
11	1376.5	32.0	928	2	092398
12	1364.5	31.7	772	2	09RB71
13	1353.5	31.5	914	2	086163
14	1329.5	30.9	936	2	092898
15	1328.5	30.9	936	2	09J442
16	1277.5	29.7	926	2	P71135
17	1090	25.4	973	2	092896
18	1090	25.4	995	2	09K2A1
19	1006	23.4	1276	2	09JRW2

20	1005	23.4	922	2	0929G5
21	1004	23.4	922	2	09K1Y9
22	996	23.2	922	2	0924B9
23	996	23.2	1407	2	092899
24	935	21.8	1013	2	084879
25	913.5	21.3	987	2	09PL45
26	790.5	18.4	867	2	09PL41
27	784.5	18.3	649	2	P71134
28	769	17.9	151	2	09J5K7
29	706.5	16.4	878	2	084882
30	617	14.4	359	2	09J5K6
31	611.5	14.2	445	2	09RB67
32	599.5	13.9	947	2	09J5E2
33	595.5	13.9	947	2	09RB70
34	593.5	13.8	427	2	09RB70
35	559	13.0	494	2	09RB68
36	554	12.9	514	2	0923D6
37	539	12.5	1672	2	0923D6
38	536.5	12.5	1723	2	092812
39	536.5	12.5	1723	2	09RB59
40	536.5	12.5	1732	2	09K2C1
41	520.5	12.1	976	2	09PL47
42	506.5	11.8	1609	2	09RB58
43	506.5	11.8	1609	2	0926U5
44	503.5	11.7	934	2	09J5E7
45	501.5	11.7	934	2	092882

## ALIGNMENTS

RESULT 1  
ID 0923A1 PRELIMINARY; PRT; 841 AA.  
AC 0923A1;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE OUTER MEMBRANE PROTEIN 7 (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).  
GN OMP7 OR PMP\_2 OR CP0761.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CHL029;  
RX MEDLINE=9920606; PubMed=10192388;  
RA Kaiman S., Mitchell W., Marathe R., Lamme J., Hyman R.W.,  
RA Klingner L., Grimwood J., Davis R.W., Stephens R.S.;  
RL "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.",  
RL Nat. Genet. 21:385-389(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VR1310;  
RA Hierne K., Boesen T., Daugaard L., Knudsen K., Madsen A.,  
RA Christiansen G., Birkelund S.;  
RT "Chlamydia proteins containing the GGAT-repeat belong to a subfamily  
of auto-transferring pathogenicity factors.",  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Unayam L.A., Ullrich T.,  
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
RA Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,  
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
pneumoniae AR39.",  
RT Nucleic Acids Res. 28:1397-1406(2000).  
RL EMBL; AB001586; AAD18172.1; -  
DR EMBL; AJ133035; CAB37083.1; -

DR EMBL: AE002235; AAF36561.1; -  
 DR TIGR: CP0761; -  
 DR InterPro: IPR003357; -  
 DR InterPro: IPR003368; -  
 DR Pfam: PF02385; OMP: 1.  
 DR Pfam: PF02415; DUF145; 1.  
 DR SEQUENCE 841 AA; 89600 MW; 9064D6D0D678D24C CRC64;

Query Match 99.5%; Score 4278; DB 2; Length 841;  
 Best Local Similarity 99.5%; Pred. No. 1.3e-244;  
 Matches 837; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKIPLRFLISLVPTLSMSNLGAATTEELASNSFDGTTSTSFSSKTSATDGTNYVF 60  
 DB 1 MKIPLRFLISLVPTLSMSNLGAATTEELASNSFDGTTSTSFSSKTSATDGTNYVF 60  
 QY 61 KDSVYIENVKKTGETOSTSCFKNDAAAGDNLFLGGGFTFFSNIDATTAAGAAIGSEAN 120  
 DB 61 KDSVYIENVKKTGETOSTSCFKNDAAAGDNLFLGGGFTFFSNIDATTAAGAAIGSEAN 120  
 QY 121 KTYTLSCFSALSLFKSPASTVTNGLGAINVKGNLSLNDNKKVLIDNFTSGDGAICAC 180  
 DB 121 KTYTLSCFSALSLFKSPASTVTNGLGAINVKGNLSLNDNKKVLIDNFTSGDGAICAC 180  
 QY 121 KTYTLSCFSALSLFKSPASTVTNGLGAINVKGNLSLNDNKKVLIDNFTSGDGAICAC 180  
 DB 121 KTYTLSCFSALSLFKSPASTVTNGLGAINVKGNLSLNDNKKVLIDNFTSGDGAICAC 180  
 QY 181 SLKIANKSLSFIGNSSSTRGCAIHTKNLTLSGGETLFOGNTAPTAAGGGAIAIDSG 240  
 DB 181 SLKIANKSLSFIGNSSSTRGCAIHTKNLTLSGGETLFOGNTAPTAAGGGAIAIDSG 240  
 QY 241 TSLISGDSGDIIEGNTIGATGVSHSIDLGTSAKITALRAAOGHTIYYPDTVTGST 300  
 DB 241 TSLISGDSGDIIEGNTIGATGVSHSIDLGTSAKITALRAAOGHTIYYPDTVTGST 300  
 QY 301 SVADALINSPDGNKKEYGTIVFSGEKLTEAARDKRTSKLONVAFKNGTVVLKG 360  
 DB 301 SVADALINSPDGNKKEYGTIVFSGEKLTEAARDKRTSKLONVAFKNGTVVLKG 360  
 QY 361 DVLVANGFQSDANSKLIMDLGTSLVANTESIELTNEINIDSLRNGKIKLSAATAOKD 420  
 DB 361 DVLVANGFQSDANSKLIMDLGTSLVANTESIELTNEINIDSLRNGKIKLSAATAOKD 420  
 QY 421 IRIDRPVVALISDESFYQNGFLNEDHSYDGIIELDAGKDIVISADSRSIDAVOSPYGYOG 480  
 DB 421 IRIDRPVVALISDESFYQNGFLNEDHSYDGIIELDAGKDIVISADSRSIDAVOSPYGYOG 480  
 QY 481 KWTINNSTDKKATVSAKOSFNPTAEQAPLVNPLMGSFIDVRPQNFIEGTGEGAPY 540  
 DB 481 KWTINNSTDKKATVSAKOSFNPTAEQAPLVNPLMGSFIDVRPQNFIEGTGEGAPY 540  
 QY 541 EKRFVWAGISNVLHRSGRENRKFRHVSAGAVGASTRMPGCDTSLGFAQLFARDKDYF 600  
 DB 541 EKRFVWAGISNVLHRSGRENRKFRHVSAGAVGASTRMPGCDTSLGFAQLFARDKDYF 600  
 QY 601 MNMNFKTYAGSLRLQDADSLYSVSLLEGGRLRELLPYVSKTLPYCSRYGOLSVGHTD 660  
 DB 601 MNMNFKTYAGSLRLQDADSLYSVSLLEGGRLRELLPYVSKTLPYCSRYGOLSVGHTD 660  
 QY 661 HRKKTSLPPPTSLTDHSMGYSWAGELGTRVAVENTSGRGFEYPPFYKQAVVYS 720  
 DB 661 HRKKTSLPPPTSLTDHSMGYSWAGELGTRVAVENTSGRGFEYPPFYKQAVVYS 720  
 QY 721 RODSFVELGASIDFSDSLILYLAIPLGKLEKRFADQYHVVAAMSPDVCRSNPKCTTT 780  
 DB 721 RODSFVELGASIDFSDSLILYLAIPLGKLEKRFADQYHVVAAMSPDVCRSNPKCTTT 780  
 QY 781 LLSNOSGSKTKGSLNARQAGIVQASGRSLGAAALFGNGCFEMRGSSRYANPDAGSKIK 840  
 DB 781 LLSNOSGSKTKGSLNARQAGIVQASGRSLGAAALFGNGCFEMRGSSRYANPDAGSKIK 840  
 QY 841 F 841  
 DB 841 F 841

RESULT 2  
 ID 09RB73 PRELIMINARY; PRT; 712 AA.  
 AC 09RB73;  
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)  
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY.  
 GN PMP 2.1.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shiba T., Ishii K., Hattori M., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RT Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RL from Japan and CWL029 from USA.  
 DR EMBL: AP002545; BAA9823.1; -  
 DR InterPro: IPR003357; -  
 DR InterPro: IPR003368; -  
 DR Pfam: PF02385; OMP: 1.  
 DR Pfam: PF02415; DUF145; 1.  
 DR SEQUENCE 712 AA; 75569 MW; 31B6A49A113E201 CRC64;

Query Match 79.0%; Score 3394; DB 2; Length 712;  
 Best Local Similarity 99.7%; Pred. No. 1.6e-192;  
 Matches 670; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKIPLRFLISLVPTLSMSNLGAATTEELASNSFDGTTSTSFSSKTSATDGTNYVF 60  
 DB 1 MKIPLRFLISLVPTLSMSNLGAATTEELASNSFDGTTSTSFSSKTSATDGTNYVF 60  
 QY 61 KDSVYIENVKKTGETOSTSCFKNDAAAGDNLFLGGGFTFFSNIDATTAAGAAIGSEAN 120  
 DB 61 KDSVYIENVKKTGETOSTSCFKNDAAAGDNLFLGGGFTFFSNIDATTAAGAAIGSEAN 120  
 QY 121 KTYTLSCFSALSLFKSPASTVTNGLGAINVKGNLSLNDNKKVLIDNFTSGDGAICAC 180  
 DB 121 KTYTLSCFSALSLFKSPASTVTNGLGAINVKGNLSLNDNKKVLIDNFTSGDGAICAC 180  
 QY 181 SLKIANKSLSFIGNSSSTRGCAIHTKNLTLSGGETLFOGNTAPTAAGGGAIAIDSG 240  
 DB 181 SLKIANKSLSFIGNSSSTRGCAIHTKNLTLSGGETLFOGNTAPTAAGGGAIAIDSG 240  
 QY 241 TSLISGDSGDIIEGNTIGATGVSHSIDLGTSAKITALRAAOGHTIYYPDTVTGST 300  
 DB 241 TSLISGDSGDIIEGNTIGATGVSHSIDLGTSAKITALRAAOGHTIYYPDTVTGST 300  
 QY 301 SVADALINSPDGNKKEYGTIVFSGEKLTEAARDKRTSKLONVAFKNGTVVLKG 360  
 DB 301 SVADALINSPDGNKKEYGTIVFSGEKLTEAARDKRTSKLONVAFKNGTVVLKG 360  
 QY 361 DVLVANGFQSDANSKLIMDLGTSLVANTESIELTNEINIDSLRNGKIKLSAATAOKD 420  
 DB 361 DVLVANGFQSDANSKLIMDLGTSLVANTESIELTNEINIDSLRNGKIKLSAATAOKD 420  
 QY 421 IRIDRPVVALISDESFYQNGFLNEDHSYDGIIELDAGKDIVISADSRSIDAVOSPYGYOG 480  
 DB 421 IRIDRPVVALISDESFYQNGFLNEDHSYDGIIELDAGKDIVISADSRSIDAVOSPYGYOG 480  
 QY 481 KWTINNSTDKKATVSAKOSFNPTAEQAPLVNPLMGSFIDVRPQNFIEGTGEGAPY 540  
 DB 481 KWTINNSTDKKATVSAKOSFNPTAEQAPLVNPLMGSFIDVRPQNFIEGTGEGAPY 540  
 QY 541 EKRFVWAGISNVLHRSGRENRKFRHVSAGAVGASTRMPGCDTSLGFAQLFARDKDYF 600  
 DB 541 EKRFVWAGISNVLHRSGRENRKFRHVSAGAVGASTRMPGCDTSLGFAQLFARDKDYF 600



```

QY 601 MNTNFAKTAGSLRLQHDASLSVSVSILLGEGGRELLPYVSKTLPCSPYCOLSYGHD 660
DB 601 MNTNFAKTAGSLRLQHDASLSVSVSILLGEGGRELLPYVSKTLPCSPYCOLSYGHD 660
QY 661 HRMKTESLPPPP 672
DB 661 HRMKTESLPPPP 672
RESULT 3
P77792 PRELIMINARY; PRT: 839 AA.
ID P77792:
AC P77792:
DT 01-FEB-1997 (TREMBLREL. 02, Created)
DE 01-FEB-1997 (TREMBLREL. 02, last sequence update)
DE 01-MAR-2001 (TREMBLREL. 16, last annotation update)
DE POMP90B PRECURSOR.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RL Submitted (Oct-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RL Submitted (Oct-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL: U65942; AACC15922.1; -.
DR EMBL: U65942; AACC15922.1; -.
DR InterPro: IPR003357; -.
DR InterPro: IPR003358; -.
DR Pfam: PF02385; OMP; 1.
DR Pfam: PF02415; DUF145; 1.
KW SIGNAL.
FT CHAIN 1 16 POTENTIAL.
FT SIGNAL 1 16 POMP90B.
SQ SEQUENCE 839 AA; 89824 MW; 4581C7CBAF7FF4C4 CRC64;
Query Match 34.8%; Score 1495; DB 2; Length 839;
Best Local Similarity 41.0%; Pred. No. 3.3e-80;
Matches 357; Conservative 132; Mismatches 321; Indels 60; Gaps 25;
QY 1 MKTPELLSLVPTLSMSNLGAATTEE--LSASNSPDGTTSTSTSPSSKTSATDGTNY 58
DB 1 MKHPVYMFLLS--SLFASNSLSFANDAOATLPPSDSYNGVNTSEEQYKETSS--GTTY 56
QY 59 VFRDSVIVNERTGTGTOSTSCPKNDAAAGDLNLFGGGSSFTFSNDATYATGAAIGSEA 118
DB 57 TCGDNVICSPAGDSGLK--SCF--SATDNLFLGNGYTLCPDNTTITASNPGAIVOG 112
QY 119 ANKTVLSEFSALSPKSPATVTEGLAIIVKGNLSLIDNDKVLIDONNSTDGGAINC 178
DB 113 QGKTLISGFSLSFCAYCPG--TTGYGALQTKGNTLIDNSSLVTHKNCSTAGGAIQC 170
QY 179 AGS---LKIANKSLSPFGNSSSTRGAIHTKNTLSSGGETLFOGNTAPTAAG--KGGA 233
DB 171 KGSDDAELEKINNQNIVFSENSSTSGAIVADKLTIVSGPFLFNNSSVNSGSSPKGGA 230
QY 234 IAIAD-SGLTSLISGSDIIFEGN---TIGATGVSHSIDLGTSAKTITALPAAGGHT 288
DB 231 ISTKDSGECSLNADADGTFDGNKIITSGGSSVTNRSIDGT--GKFTKLAKGFGI 289
QY 289 YFVDPTTVGSTVADALINSPDGNKEVGTIVSGSKTLFEAKDEKNTSKLON 348
DB 290 FFYDPTITGGG---DELNINKETVD--YTKIVFSGEKLDEKARAKENASTFNP 342
QY 349 VAFKNGTVLKGVDVLSANFSDANSKILMDGTSL---VANTESIELTLEINIDSLR 405

```

```

DB 343 ITLSAGSLVLDKGVSVYAKQVQEGAGTVYMDLGTTLQTPSSGGETITLTLNDINIASLG 402
QY 406 NG---KKILSAATQAKDRIKRPVYLAISSDEFYONGPLNEBHSIDGILELDAGDIY 462
DB 403 GGGGTSAPKATNTATASOATTIN--AVNLVDADGNAVYDPLATSKPPTALVATTNASTVQ 461
QY 463 SADSRSINAVOSPYYGKGTINMSTD--DKKATVSAKOSFPPTAEAPLVPNLIMS 520
DB 462 PPDNLNTNVPPTHYQGNMTYVMDTETATKTATLTWEOTGSPNPERGQPLVPNLKMA 521
QY 521 FIDVRPQNFIELGTEGAPYKRRFVAGISNVLHRSRENQRRFVSGAVVASTRMP 580
DB 522 FSDLRALQVMDISVNGADYHNGFVWGLANFLHKSQSPDKRFRNSAGYALGVAKTP 581
QY 581 GGDITLSLGFQALFARQKDYFMNTNFAKTAGSLRLQHDASLSVSVSILLGEGGRELLP 640
DB 582 SDDIFSAACQQLFGKDKDYLVSNNBNITAGSLYXOH--ISYWS-----AMQNLON 631
QY 641 YVSKTLPCSPYCOLSYGHDHMKTE---SLPPPPPTLSMDHTSMGYYVAGELGTRVAV 697
DB 632 TIGAEPPLVLAQVLTCHASNDKMTNTTTPYAPKRTTYAEIKDQKNGDGVELGATVP 691
QY 698 ENTSGRFFREYTPPVKYAVYSDROSEVELGA--ISRDSDSLNYLAIPLGILEKRA 756
DB 692 QTESSSL-LFDWVSPFLKFLVHTHODDFKENSQDGRFESSNLPLSLIPIGIKFE-RFA 749
QY 757 EQ---YHYVAVMTSPVCKSNPKCTTTLISNQS--WKTKGSLAQAQGVQASGRSLG 811
DB 750 NNDTASYHYTAAYSPIVRSNPDCTTSLVSPSAVWVKANNIARSAPFALOGNYLIS 809
QY 812 AAELFGNFGEWRGSSRSYVNAVSGKIF 841
DB 810 HNTIFSQFEELRGSSRYTVNDLGSRIOP 839
RESULT 4
P71132 PRELIMINARY; PRT: 847 AA.
ID P71132:
AC P71132:
DT 01-FEB-1997 (TREMBLREL. 02, Created)
DE 01-FEB-1997 (TREMBLREL. 02, last sequence update)
DE 01-MAR-2001 (TREMBLREL. 16, last annotation update)
DE POMP91A.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RA MEDLINE-96406378; PubMed-8810511;
RL Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RT "Identification of a multigene family coding for the 90 kDa proteins of the ovine abortion subtype of Chlamydia psittaci.";
RT FEMS Microbiol. Lett. 142:277-281(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RA MEDLINE-98187897; PubMed-9529048;
RL Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RT "Molecular cloning and characterization of the genes coding for the highly immunogenic cluster of 90-kilodalton envelope proteins from the Chlamydia psittaci subtype that causes abortion in sheep.";
RL Infect. Immun. 66:1317-1324(1998).
DR EMBL: U65942; AACC15921.1; -.
DR InterPro: IPR003357; -.
DR InterPro: IPR003358; -.
DR Pfam: PF02385; OMP; 1.
DR Pfam: PF02415; DUF145; 1.
SQ SEQUENCE 847 AA; 90694 MW; 754C958E7F1179E CRC64;
Query Match 34.5%; Score 1482; DB 2; Length 847;
Best Local Similarity 39.4%; Pred. No. 2e-79;

```

Matches 346; Conservative 147; Mismatches 316; Indels 70; Gaps 24;

```

OY 1 MKIPLELLISLVPLTSMNLGAATTEE--LSASNSFDGTTSTFSRSKTSATDGTNY 58
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKRPVWFLIS--SLIVSNLSXSEEDOKTLTSAHSYNNTNTSEPPNPLSTNSNCTIY 58
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 59 VFKDSVVIENPVKTGTOSTCFKNDAAAGDLNPLGGSFTFSNIDATTASGAIGSEA 118
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 59 TCTGNICIAVAGLDGSLSSSCFTD--TAGNLSFLNGNYLFCGDNITWTOSSHGAISVSG 116
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 119 ANKTVTLGSFASLFLKSPASTVINGLATNVAGNLSLNDNDVLIQDNSTGDDGAINC 178
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 117 TMTTLISGSLFSCAYCPG--ATGGAIKAVGNTTITKDNSSLVVRHKNSTEGGAIQC 174
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 179 AGS-----LKIANKSLSFIGNSSSTRGGAIHKNLTLSGGETLFOGANTAPPAAGKGA 233
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 175 KASSSEAELEKIENNOMLVPAENSSSGGATVADKLTIVSGPTLSSNNSVSAKSGGA 234
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 234 IAIADS-GTILISGSDITFEBCNTI---GATGIVSHSIDLGTSAKITLAPAAOGHTT 288
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 235 ICIKDGGECSTLADLGDITFDGNKIITKNGGSPFYVRNSIDGSSGKFTKLAKGEGFI 294
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 289 YFPDPTTVGTSVADALINSPDGNKEKTYTIVSGEKLFEAKDKNNTSKLQN 348
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 295 FYDPTITGGG---DELNINKODIYD--YTKIVSGERLSEDEKKAANKKSKPDKOP 347
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 349 VAEKNGTVYLGKDVLSANGFSODANSKLMDLGTSL---VANTESELTNLEINIDSLR 405
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 348 LKIGSSLLIKDGVLTLETFSFTQTEGATVMDLGTLLQPPSSGGETITLTLNIDVASSIG 407
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 406 NG-----KKIKLSAATKOKIRIDRPVYLAISDSFYQNGFLNDSHYDGLIEDACKDIY 461
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 408 GGGVADDPKAVETTESKVTIN-AVNLVDNCAEYPLAASQEPFAL-EVRSSSGS 465
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 462 ISADSRISINAVOSP--YGYGKWTINW---STDDKKAATVAKOSFNPAPAEADLVPR 515
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 466 ITRPTTNLENYTPPTHTGYGNTVTWKOGSSAKQEKTAITLMEOTGSPNPERKOSLVRN 525
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 516 LMGSEFIDVRPFONTELEGEAPYERKRFVAVAGISVNLHRSRENORRFRHVSAGAVGA 575
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 526 TLMGSESDIRAIQNLMDISVNGADYHRGFWSGLGNFLHKSQSDTRKFRHNSAGVALGY 585
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 576 STRMGCGDLSLGFQALFKRDKDYFNNTNPAKTYAGSLRLOHDASLYSVASITLGGGGLR 635
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 586 YAOPTSEDEYSAFCOLFEKDKDYLVSKNSSYVYAGSIYQH-ISYWNMTWNTLL----- 638
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 636 ELLLPVYSKTL---PCSPFYGOLSYGHTDHRMK---TESLPPPPPLSLDHTSMGUYMA 688
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 639 -----QNTLGAELAPLVLAQALYCHASNMMKNTMTDYPAPKTTYSEIKGDWMDCFG 691
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 689 GELGTRVAVENTSGRGCFREYTPFVKQAVYVSRODSFVELGA-ISRDFSDSHLYNLAIP 747
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 692 VEEGAKAPIETAS--LLFWMYSPFYQLVVAHQODFKNNNSDQGYFESNNLTJNSMFI 749
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 748 GIKLEKREAEQ---YHYVAMTSPDYCRSNPKCTITLNSNGS--WKTGSLNARQAGIV 802
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 750 GYKLEK-FSHKKTASNLNLAAPDIVRSNPCTASLIVSPTSAAVWYTKANNLARAAFTL 808
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 803 QASGFRSLAAALEFCNFGFEWGSRSRYNVDAGSKIRF 841
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 809 QAGNTIALTRNTELSQGFELRGSCRTYNIDLSKIQF 847
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. NCBI\_TaxID=83558;

OX [1]

RN SEQUENCE FROM N.A.

RP STRAIN-CDC/WR-029/VR-1310;

RC Knudsen K., Madsen A.S., Myding P., Christiansen G., Birkegaard S.;

RA "Identification of two novel genes encoding outer membrane complex

RT associated surface layer proteins in Chlamydia pneumoniae.";

RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-VR1310;

RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,

RC Christiansen G., Birkegaard S.;

RT "Chlamydia proteins containing the GCAT-repeat belong to a subfamily

RT of autotransporting pathogenicity factors.";

RL Submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-CWL029;

RA MEDLINE=99206606; PubMed=10192388;

RC Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,

RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;

RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";

RL Nat. Genet. 21:385-389(1999).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN-J138;

RX MEDLINE=20330349; PubMed=10871362;

RA Shirai M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,

RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;

RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138

RT from Japan and CWL029 from USA.";

RL Nucleic Acids Res. 28:2311-2314(2000).

DR EMBL: AJ001311; CA04672.1; -

DR EMBL: AJ133034; CA037072.1; -

DR EMBL: AE001628; AAD18593.1; -

DR EMBL: AF002546; BAA98658.1; -

DR InterPro: IPR003357; -

DR InterPro: IPR003368; -

DR Pfam: PF02385; OMP; 1.

DR Pfam: PF02415; DUF145; 1.

KW Signal.

FT CHAIN 1 17 POTENTIAL.

FT SIGNAL 18 928 OUTER MEMBRANE PROTEIN 4.

FT SEQUENCE 928 AA; 98903 MW; 788BCDD62C911402 CRC64;

Query Match 33.3%; Score 1429.5; DB 2; Length 928;

Best local Similarity 35.8%; Pred. No. 2.9e-76;

Matches 339; Conservative 147; Mismatches 336; Indels 125; Gaps 15;

```

OY 1 MKIPLELLISLVPLTSMNLGAATTEELASNSFDGTTSTFSRSKTSATDGTNY 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKTSLVWLVSSVLAAS-CHLOSLANEELISPDSPNGINDSTFTPKISA-----TLYSL 55
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 61 KDSVVIENPVKTGTOSTCFKNDAAAGDLNPLGGSFTFSNIDATTASGAIGSEAN 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 56 TGVDFEYE-PGKGTPLSDSCFKO--TTDNLTFELNGHSLTFEGIDAGTAGAA-ASTTAN 111
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 121 KVTYLSGFALSFLKSPASTVINGLATNVAGNLSLNDNDVLIQDNSTGDDGAINCAG 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 112 KNLTFSGFSLSPDSSPSTVTGOGTLLSAGGVLENIRKLIVAGNFSTADGAIKGS 171
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 181 ----- 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 172 FLITGSGDALSPNNSSSTKGAIAATTAGARIANNNTGYRFLSNIASTSGAIDDEGTSI 231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 181 -----SLKIANKSLSFIGNSSSTRGGAIHKNLT 210
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 232 LANNKFLYEGNAAKTTGGAICNTKASGSPELLISNNKTLIVASVNAETSGGAHAKKLA 291
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 211 LSSGGETLFOGNTAPPAAGKGAIAIADSGTISIGSDGDIIFEGNTI---GATGTVSHS 267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 292 LSSGFEFLNNVSSATPKGAISIDASGELSLSAEAGNTTFVRNLTGTTGSDTDPKRN 351
Qy 268 AIDGTSKITALAAAGHTTYFYDPTTVGTSVADALINSDTDGNKTYETIYFSG 327
Db 352 AINISNGKFTLAAAHNTTFEYDPTTSECTSS--DVLKINNSAGALNPEYQGTITLFG 409
Qy 328 EKLAEAKDEKNTSKLLOWAVFKNGTVYKGDVYLSANGFSODANSKLIMDGTSLVA 387
Db 410 ETLADELKAADNKKSFOTOPVSLSGKLLQKGYTLSTFSQDASLLGMDGTTLTST 469
Qy 388 NTSEIETLNEINISLNRNGKIKLSATAOKDIRIDRPVYLAISDSFYONGFLNEDHS 447
Db 470 TAGSITITNLGINVDSGLKQPVSLFAKGAENKIYSGKLNLIIDEGNITESHMFSDQL 529
Qy 448 YDGLIEFDAGKDIYISADSINSINAV-----QSPYIGOKKWTINSTD---DKATVSWA 498
Db 530 F-SLTKITVDADVDTNDISLIPVPAEDPNSEYGFQGNVNMVTTDTATNTKEATATWT 588
Qy 499 KQSNPTAEDEAPLVNPLMGSFIDVPRPFQNFIELGEGAPYEKRFVWAGISNVLHSGR 558
Db 589 KTGVPSPERSKALVCNTLMGVFTDIRSLQOLVEIGATGMEHKGFEVSSMTNFLHKTGD 648
Qy 559 ENQKRFHVSGGAVVAGSTRMPGCDTSLGFAQLFARDKDYFMNTNFAKTYAGSLRLQHD 618
Db 649 ENRKGFPHTSGGYIGGSAHTRPKDDLTFAFCHLFARDKOCFIHNNSRVYGTGLFFKHS 708
Qy 619 ASLYSVSILGEGGLREILIPYVSKTLPCSFYQOLSYGTHDHAKTE--SLPPPTLS 676
Db 709 HTLOPONYLRIGRAKFSESAIEKPREIPLALDVQVSFSHSDNMEHYTSLP----- 761
Qy 677 TDHNSMGVYVAGELGTRVAVENTSGRFPFRETPEPVKQVAVYRSDSFVGLAISRDES 736
Db 762 ESEGSWSNECTAGIGLDLPVLSNPHPLFTFTIPQKMEVAVYVSQNSFFSSSDGKGS 821
Qy 737 DSHLYNLAIPLGIG-LEKRFAPQY-YHVVAYSBVCRSNPKCTTLLISNOGSKRTGNS 794
Db 822 IGRLLNLSIPGAKFYOGDIDGSDTYDLSGFFVSDVYRNNQSTATILVMSDMSKIRGN 881
Qy 795 LARAGIVQASGFRSLGAAAELEFNGFEMRGSSRSRYNDAGSKIRF 841
Db 882 LSRQAFILRGSNVYVNSNCELFGHYAMELGRSSRYNVDVGTKLRF 928

RESULT 6
QyK299 PRELIMINARY; PRT; 949 AA.
AC QyK299;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
GN CP0302.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RA MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Unayam L.A., Uterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khoriti H., Craven B.,
RA Bowman C., Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AE002192; AF38159.1; -.
DR TIGR; CP0302; -.
DR InterPro; IPR003357; -.
DR InterPro; IPR003368; -.
DR Pfam; PF02385; OMP; 1.

```

```

DR Pfam; PF02415; DUF145; 1.
SQ SEQUENCE 949 AA; 101357 MW; A00B09E16C69BE3 CRC64;
Query Match 33.3%; Score 1429.5; DB 2; Length 949;
Best Local Similarity 35.8%; Pred. No. 3e-76;
Matches 339; Conservative 147; Mismatches 336; Indels 125; Gaps 15;

```

```

Qy 1 MKIPLRFLISLIVPTLMSNLGAATTEELASNSPGTSTSTSSKTSATGCTNYVF 60
Db 22 MKTISIPVAVSVLAFS-CHLOSIANELELSPDSEFGNIDSGFTFKTSA---TYSL 76
Qy 61 KDSVINEVPRKTEGTSQSCFEKNDAAAGDLNFLGCGSFTEFSNIDATTASGAIGSPAAN 120
Db 77 TGDVFEYE-FKGTGPLDSCRKQ--TTDNLTFLGNGSLTFGFLDAGTHGAA-ASTTAN 132
Qy 121 KTVTLGFSALSFLKSPASTVINGLAINKGNLSLNDKVLIQDNESTGDGAINCAG 180
Db 133 KNLTFSGFSLISFDSSPTVYTGQGLSSAGVNLNIRKLVYAGNFSFADGAIKQAS 192
Qy 181 ----- 180
Db 193 FLITGTSGLALFNSNSSSTRKGAIAATTAGARIANNNGYRELINIASTSGAIDEGTSI 252
Qy 181 -----SLKIANKSLSFIGNSSSTRGAIHTKMLT 210
Db 253 LSNKKFLYFEGNAKKTGGAICNTRKAGSPELLISNKKTLIFASNVAETSGAIIHAKKIA 312
Qy 211 LSSGGETLFGONTAPTAAKGAIAIADSGTLLISGDSGDIIFEGNTI--GATGVSHS 267
Db 313 LSSGFEFLNNVSSATPKGAISIDASGELSLSAEAGNTTFVRNLTGTTGSDTDPKRN 372
Qy 268 AIDGTSKITALAAAGHTTYFYDPTTVGTSVADALINSDTDGNKTYETIYFSG 327
Db 373 AINISNGKFTLAAAHNTTFEYDPTTSECTSS--DVLKINNSAGALNPEYQGTITLFG 430
Qy 328 EKLAEAKDEKNTSKLLOWAVFKNGTVYKGDVYLSANGFSODANSKLIMDGTSLVA 387
Db 431 ETLADELKAADNKKSFOTOPVSLSGKLLQKGYTLSTFSQDASLLGMDGTTLTST 490
Qy 388 NTSEIETLNEINISLNRNGKIKLSATAOKDIRIDRPVYLAISDSFYONGFLNEDHS 447
Db 491 TAGSITITNLGINVDSGLKQPVSLFAKGAENKIYSGKLNLIIDEGNITESHMFSDQL 550
Qy 448 YDGLIEFDAGKDIYISADSINSINAV-----QSPYIGOKKWTINSTD---DKATVSWA 498
Db 551 F-SLTKITVDADVDTNDISLIPVPAEDPNSEYGFQGNVNMVTTDTATNTKEATATWT 609
Qy 499 KQSNPTAEDEAPLVNPLMGSFIDVPRPFQNFIELGEGAPYEKRFVWAGISNVLHSGR 558
Db 610 KTGVPSPERSKALVCNTLMGVFTDIRSLQOLVEIGATGMEHKGFEVSSMTNFLHKTGD 669
Qy 559 ENQKRFHVSGGAVVAGSTRMPGCDTSLGFAQLFARDKDYFMNTNFAKTYAGSLRLQHD 618
Db 670 ENRKGFPHTSGGYIGGSAHTRPKDDLTFAFCHLFARDKOCFIHNNSRVYGTGLFFKHS 729
Qy 619 ASLYSVSILGEGGLREILIPYVSKTLPCSFYQOLSYGTHDHAKTE--SLPPPTLS 676
Db 730 HTLOPONYLRIGRAKFSESAIEKPREIPLALDVQVSFSHSDNMEHYTSLP----- 782
Qy 677 TDHNSMGVYVAGELGTRVAVENTSGRFPFRETPEPVKQVAVYRSDSFVGLAISRDES 736
Db 783 ESEGSWSNECTAGIGLDLPVLSNPHPLFTFTIPQKMEVAVYVSQNSFFSSSDGKGS 842
Qy 737 DSHLYNLAIPLGIG-LEKRFAPQY-YHVVAYSBVCRSNPKCTTLLISNOGSKRTGNS 794
Db 843 IGRLLNLSIPGAKFYOGDIDGSDTYDLSGFFVSDVYRNNQSTATILVMSDMSKIRGN 902
Qy 795 LARAGIVQASGFRSLGAAAELEFNGFEMRGSSRSRYNDAGSKIRF 841
Db 903 LSRQAFILRGSNVYVNSNCELFGHYAMELGRSSRYNVDVGTKLRF 949

```



DR InterPro: IPR003357; -  
DR InterPro: IPR003368; -  
DR Pfam: PF02385; OMP; 1.  
DR Pfam: PF02415; DUF145; 1  
DR SEQUENCE 930 AA; 97639 MW; 46A8896761391C09 CRC64;

Query Match 32.9%; Score 1413.5; DB 2; Length 930;  
Best Local Similarity 37.0%; Pred. No. 2.5e-75;  
Matches 355; Conservative 124; Mismatches 332; Indels 149; Gaps 20;

```
OY 1 MKIPLRLILSLVPTLSMSNLGAAT---TEELASNSFSDGTSTTSFSSKTSATDGTN 57
DB 1 MKIPLRLILSLVPTLSMSNLGAAT---TEELASNSFSDGTSTTSFSSKTSATDGTN 57
OY 58 YVRKDSYVLENPKTGETOSTSCFKNDAAADINFLGGGFTSPFSNIDATTASGAAGSE 117
DB 58 YVRKDSYVLENPKTGETOSTSCFKNDAAADINFLGGGFTSPFSNIDATTASGAAGSE 117
OY 118 AAKRTYLSGFSALSFKSPASTVTNGLGAINVGNLSLIDNKKVLIQDNFS----- 169
DB 114 TADKALFTGFSNLSFTIAPGCTIVASGKSTLSAGALINLIDNCTILFSQVNSNEANNNG 173
OY 170 ----- 169
DB 174 AITAKTLISGNTSITFTSNSAKRLGAIYSGAASISGNTGQLVEMNNKGETGGALG 233
OY 170 -----TGDCGAINC-----AGSLKATANKSLSTFIONSSSTRG 201
DB 234 FEASSSITONSLFSGGNATDAAGKGAALYCKRTGETPPLTISGKSLFAENSSVTQG 293
OY 202 GAHTKNTLTLSSGCTLFQGN-TAPPAAGKGAIALADSGTSLISGDSGDIIEGNWIGA 260
DB 294 GATCAGGLDLSAAGPILFSSNRCGNTAAGKGAIALADSSLSLSANOGDITLGLNTLS 353
OY 261 TG--TVSHSAIDIGTSAKITATLAAOGHTIYFYDPITVTGTSVADALINSPDGNKE 318
DB 354 TSAPTSTRNALYIGSSAKITNLRAAGOSIYFYDPI-ASNTTGASDVLITINQDSNPLD 412
OY 319 YRTIYFSSGKLFLEAKAKDEKNTSKLQNVAKNGIVYKAGVYLSANGFSDANSKLI 378
DB 413 YSGTIYFSSGKLSADAKADNFTSLIKOPALASGLAKGVEIDVNGFTQEGSTL 472
OY 379 MDGTSVAVNTESIELTNEINIDSLRNGKKIKLSATAOKDIRIPRYLAISDESFYQ 438
DB 473 MGTGTLKADTEISLTKLVYDLSALEGNKSVSIEFAGAKTITLSPLEFQDSGGEYE 532
OY 439 NGFLNEDHSYDIL--ELDAGKDIVISADSRSINAVOSP---YGOGKWTINN--STDDK 491
DB 533 SHTINQAFQPLVFTAAATAAADLYI--DALITSPTQPEPHYGCGHWEATWADISTAK 590
OY 492 KATVSAKAKOSFNTPAQEAPLVNLLMGSFIDVRPFQNFELTEGAPYEKKRWAGISN 551
DB 591 SGTMTVTTGYNENPNRRASVVPDLSMASPTDIRLQOIMTSQANSIYQGRGLMASGTAN 650
OY 552 VLRSGRENOKRRAHSGAVVCASTRMPCGDTLSLGFALDFAKRDYFAMNTFAKTYAG 611
DB 651 EFKKDSGCTNOAFRRKHSYGIYVGSADBSSENFVAFCQLFKKDKDLFVEMTSHNYLA 710
OY 612 SLRLQDASLYSVVSTILLEGGLRETLIPYVS-----KTLPCSFYQGLSYGHTDHRM 663
DB 711 SLYLQRRALF-----GG-----LPMPSRSGITMDLKIDIPILINQGLSYSTRKNDM 755
OY 664 KTSLSLPPPTLSTDBTSMNGYVWAGELGTRAVENTSGRGFRETTPYKVAIVSRQD 723
DB 756 DFTYTSYPEA-----OGSMNNSGALLEGSLALYLPKPAEPFQGYFPPLFKPAVYSRQ 810
OY 724 SFYELGAISDFSDSHYNLAIPLGIKLEKFAEO--YYHVAMYSBDVCRSPKCTTTL 761
DB 811 NFESGAEAFADGDGLVNCISIPVGRLEKISDEKKNNEISLATYIGDYRKRPRTSL 870
OY 782 LSNQSGKTKGSMILARQAGIVQAGSFRSLGAAALEFGNCFEWRGSSRSYNDVAGSKIKP 841
```

```
DB 871 MVSASWTSILCKNLARQAFILASAGSHLLTSPHVELSGEAEYELGSAHIYVDCGLRYSE 930
RESULT 9
ID 09RB65 PRELIMINARY; PRT; 928 AA.
AC 09RB65; 09RB64; 09S6P2;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (OUTER MEMBRANE PROTEIN 5
DE PRECURSOR) (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
PMP.10 OR OMP5 OR CP0303.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83358;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE=20330349; PubMed=10671362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CwL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-VK1310;
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birke Lund S.;
RT "Chlamydia proteins containing the GCAT-repeat belong to a subfamily
RT of autoantigenic pathogenicity factors."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN-AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterberg T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwinn M., Nelson W., Debby R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AP002546; BA98657.1; -
DR EMBL; AJ133034; CAB37071.1; -
DR EMBL; AE002192; AAF38160.1; -
DR TIGR; CP0303; -
DR InterPro: IPR003357; -
DR Pfam: PF02385; OMP; 1.
KW Signal.
SQ SEQUENCE 928 AA; 97229 MW; 0590D5206A1D0E1 CRC64;
```

Query Match 32.6%; Score 1400.5; DB 2; Length 928;  
Best Local Similarity 34.3%; Pred. No. 1.5e-74;  
Matches 330; Conservative 161; Mismatches 315; Indels 157; Gaps 23;

```
OY 1 MKIPLRLILSLVPTLSMSNLGAATTEELASNSFSDGTSTTSFSSKTSATDGTNYV 59
DB 1 MKSQFMSLVLSLTLACTSGSVFAATAENIGPSDSFGSTNGTYPK--NNTTGIDYT 58
OY 60 FKDSYVLENPKTGETOSTS--CFKNDAAADINFLGGGFTSPFSNIDATTASGAAGSE 117
DB 59 LMGDITLQNL---GDSALTLKGCFSPTTES--LSFAGKGYLSFLNI-KSSABGAAL-SV 111
OY 118 AAKRTYLSGFSALSFKSPASTVT-----NGL----- 145
DB 112 TYDKNLSLTFGSSLTFALAPSIVITTPSGKAVKCGGDLTFDNNNGITLLEKQDYCEENGCA 171
OY 146 -----GA 147
```

```

Db 172 ISTKNLSLNKSTGISSEGNKSSATGKGAICATGTVDTNTNTATPTLFSNNIAEAGA 231
Qy 148 INKGNLSLIDNKNVLIQDNFT--GDGAIACAGSLKATANKKSLFIGNSSSTGGAT 204
Db 232 INSTGCTTGTNTSLVSESVATATAGNGALSGADVITSGNQSTFSGNQVANGAT 291
Qy 205 HKNKLTLSGCG-----ENLFQGNAPTAPAGKGAIAIADSGTSLISGSDIIEGNT 257
Db 292 YAKKLTLAGSGGGGSIIFSNNIYOG-----TTAGNGAISTILAECSLSAAGIITNGNA 347
Qy 258 IGAT--GIYSHSAIDIGTSKITALPAAGCHTITFPDPTITVTSVADALINSPTGCD 315
Db 348 IYATFTPTKRNISIDIGSTAKITNLRAISGHSIFFPDPTITANTAASTDTLNLKADAGN 407
Qy 316 NKEVGTIYFSGEKLAEAKDEKNTSKLQVAFKNGVVLKGDVLSANGFESODANS 375
Db 408 SDYSGSIYFSGEKLSEDEKAVADNLTSTLKQVYTLTAGMLVKRGVTLDTKFTOTAGS 467
Qy 376 KLIMDGTSLVANTESIELTNLEINISLNGKRIKLSATAOKDIRIDRPVLAISDES 435
Db 468 SVIMDAGTTLKASTEVEVTLGLSTIPVDSLGEKKVYIAASAASKVVALSGPILLDNOGN 527
Qy 436 FVONGGLNEDHSDIGLELDAGKDIYISADSRINAVQSP--YGYGKWTINSTD---- 489
Db 528 AYENHDIGKTODF-SFVOJSA--LGTATTTDVPVAVPTVAPPTHYGYGTGWTGMDTAST 585
Qy 490 --DKKATVSAKQSENPTEAEAPLVNLLMGSEFIDVRPQNFIELTGEGAPYKRPFWA 547
Db 586 PKTKTATTLMTNTGYPLEPNEGGLPVNSLMGFSFDIOAIQVIESALTLCSDRGFWAA 645
Qy 548 GTSNVLHRSRGRENQKFRHVSAGAVVAGASTRMBGDTLISGFALPARDPDMANTFEAK 607
Db 646 GYANFLDKOKKCKRKYRHKSGGYAIGAAGQCSSENLISPAFCQLGSDKDFLVANKHMD 705
Qy 608 TVAGSLRLQHDASLYVSVSLLEGGLRELLPYVSKTLPSCSYGQSLSYGHTDHRKTES 667
Db 706 TVAGAYIYIHTIECSGFICGLLDK-----LPGSMHKLVLLEGQLAYSHVSNDRKTX 758
Qy 668 LPPPTLTSTDHTSMGCVWAGELGTRVAVENTSGRGF-----PREYFPFVVOAVYSR 721
Db 759 TAYPEV-----KSGMGNNAFNMALG-----ASSHSYPEYLHCFTYAPYIKLNLTYIR 806
Qy 722 QDSFVELGAISRDESDHLYNALPLGILKEKRA--EQYHVAVMSPDVCRSNPCKT 778
Db 807 QDSFSEKGTGEGRSFDSNLFNLPLIGVFEK-FSDCNDFSXYDLTSLYVDLIRNDPKCT 865
Qy 779 TLLLSQSGMKTKGSLAQAQIVYASGFRSLGAALNCEFEKRGSSRSYNNVADGSK 838
Db 866 TALVYISGASMETYANNLAQALQVRAGSHYASPMFEVLGQFVEVGRSSRTIYNDLGK 925
Qy 839 IKF 841
Db 926 PQF 928

```

```

RL FEWS Microbiol. Lett. 142:277-281(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE 526/3;
RX MEDLINE-98187897; PubMed-9529048;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RT "Molecular cloning and characterization of the genes coding for the
RT highly immunogenic cluster of 90-kilodalton envelope proteins from the
RT Chlamydia psittaci subtype that causes abortion in sheep.";
RL Infect. Immun. 66:1317-1324(1998).
DR EMBL; U65943; AAC15923.1; -.
DR InterPro; IPR003357; -.
DR InterPro; IPR003368; -.
DR Pfam; PF02385; OMP; 1.
DR Pfam; PF02415; DUF145; 1.
KW Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 846 POM91B.
SQ SEQUENCE 846 AA; 90834 MW; 4CDC31DC03C2964E CRC64;

Query Match 32.1%; Score 1381.5; DB 2; Length 846;
Best Local Similarity 38.5%; Pred. No. 1.7e-73;
Matches 338; Conservative 138; Mismatches 332; Indels 71; Gaps 26;

Qy 1 MKIPLELITSLVPTLSMSNLGAA--TTEELASNSFDGTSTFSKTSATDGTNY 58
Db 1 MKHPVWFLIS--SLLASMSLSFAQVTNLTLSOSYNKNTVSDSEYK--ETSGAITY 56
Qy 59 VFQDSVVIENVPKGTGTOSTGCFKNDAAQDLNLFLLGGESFYSNIDATYASCAIASEA 118
Db 57 TCEGNVICISYAGKDPL--NKSCFSE--TTENLISFINGYLCPDNTTOSHPGAISVSG 113
Qy 119 ANKTVTLGFSNALS--ELKSPATVINGAIIVKGNLSLINDKYLIDNISTGGAGAI 176
Db 114 TTKTLDISGFSLESCAYCCPGTT--GGAIGTKKTTLLKDNSSLVFHKNCSTAGAGAI 170
Qy 177 NCAGS-----LKIANKSLISFIGNSSSTRGGAIHFNLTLSGGETLFGQNT--APTAAG 230
Db 171 QCKSSSTAEKLEKNNKLVFSSENSKEKGAIYADKLRTVSGGPTLFSNVSYSHNSPK 230
Qy 231 GGAIAIADS--GTLISGDSGDIIEGNTI--GATGYSHSAIDIGTSKITALPAAG 285
Db 231 GGAICTKDSGECSLRANLADITFDGNKIITNGGSPRYTRNSIDIGSGKFTKLNKAGG 290
Qy 286 HTIYFPDITVGTSTVADALINSPTDGNKEYGTIYFSGEKLTEAEKDEKNTSKL 345
Db 291 FGIFFYDPIANTGGSTEIE-LNKTESDT--TYTKIYFSGEKLDEKTYPAANKSYF 345
Qy 346 LQNVAFKNGTVVLKGVLSANGFESODANSKLIMDGLTSL--VANTESIELTNLEINID 402
Db 346 KQPLKIGAGSLVKDGVTLFAKKITQTKGSTVYMDLGTTLQTPSSGGEYTLNLINDINIA 405
Qy 403 SLANG--KKIKLSAATAOKDIRIDRPVLAISDESEFYONGFLENHSDGLDELADGKD 459
Db 406 SLGGGGGTAPAKLATNTAASOALSI--AAVNLVNTDSWTFYEDPILISAKSFSALATTSSST 464
Qy 460 IYISADSRISINAVQSPGYGKWTINM--STDDKATVSAKQSENPTEAEQADLVPN 515
Db 465 VTPPEFNLNKNTYPTHYGYGNMTVTWKOGSSAOEKTATLTWQGTQISNPERVGSLVNR 524
Qy 516 LMGSEIDVRRPQNFIELTGEGAPYKRFVWAGISVNLHRSRGRENQKFRHVSAGAVGA 575
Db 525 TLMGASVSDIRALQNLMDISVNGADYGRGFWVSLANFLNKSNGSDTKRKRFRHSHGALGV 584
Qy 576 STRMPCGDTLSLGFADLPARDKYFMNTNPAKTYASLSRLQHDASLSYVSLILGEGGLR 635
Db 585 YQTPSDVDVCSAFCQLFKDKDYFVSKNSSTIYASIIYQH--ISYWNWNTLL----- 637
Qy 636 EILLPVVSKTL--PCSPYGOISYGHTDHRMK--TESLPPPTPLSTDHTSMGCVYWA 688
Db 638 -----QNTLGAELPLVLAQTLTYCHASNNKTKMTNTYTPRANVTPELKGDMGDCFG 690

```

```

RESULT 10
P71133 PRELIMINARY: PRT: 846 AA.
ID P71133
AC P71133:
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, last sequence update)
DE 01-MAR-2001 (TREMblrel. 16, last annotation update)
DE POM91B PRECURSOR.
OS Chlamydia psittaci (Chlamydia psittaci).
OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_Taxid=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE 526/3;
RX MEDLINE-96406378; PubMed-8810511;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RT "Identification of a multigene family coding for the 90 kDa proteins
of the ovine abortion subtype of Chlamydia psittaci.";
```

Qy	669	GELGRVRAVENTSGRGFFRETPPYKVQAAVYSRQDSPEVLGA-ISRDFSDHLYNLAIPL	74
Db	691	VEFGAKADIETAS--LFDMSIPPVKIQLYVAHDDKRENNSDGGRFFESNLINLSMPI	748
Qy	748	GKLEKREPAE---YYHVAVMSPDVCRSKRCCTTLLSNQGS--WRTKGSNLARQAGIV	802
Db	749	GVKLEK-FSKHDFASYNLTLTAYADIVRNPDCETASILLVSPTSAMVWTAKANNLARHAFIL	802
Qy	803	QASGFRSLGAAAEELFGNFGFFWRSSSSSYNDASCKIFE	841
Db	808	QAGNYLATRNTELFESQFELRLRSCRTYNIDLSGKIQF	846
RESULT	11		
ID	Q92398	PRELIMINARY:	PRT: 928 AA.
AC	Q92398;		
DT	01-MAY-1999 (TREMBLrel. 10, Created)		
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE	OUTER MEMBRANE PROTEIN PRECURSOR (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).		
CN	OMP10 OR PMP_9 OR CP0306.		
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).		
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
OX	NCBI_TaxID=83558;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CML029; Pubmed=10192388;		
RX	MEDLINE=99206606; Pubmed=10192388;		
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,		
RT	Ollinger L., Grimwood J., Davis R.W., Stephens R.S.;		
RL	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";		
RN	Nat. Genet. 21:385-389(1999).		
RP	[2]		
RC	SEQUENCE FROM N.A.		
RX	STRAIN-VRJ310;		
RA	Hierno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,		
RT	Christiansen G., Birkelund S.;		
RL	"Chlamydia proteins containing the GGAI-repeat belong to a subfamily		
RP	of autotransporting pathogenicity factors.";		
RN	submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.		
RP	[3]		
RC	SEQUENCE FROM N.A.		
RX	STRAIN-J138;		
RA	MEDLINE=20330349; Pubmed=10871362;		
RT	Shiral M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,		
RL	Shiba T., Ishii K., Hattori M., Kubura S., Nakazawa T.;		
RP	"Comparison of whole genome sequences of Chlamydia pneumoniae J138		
RN	from Japan and CML029 from USA.";		
RP	[4]		
RC	Nucleic Acids Res. 28:2311-2314(2000).		
RX	SEQUENCE FROM N.A.		
RA	STRAIN-AR39;		
RT	MEDLINE=20150255; Pubmed=10684935;		
RL	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,		
RA	White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,		
RT	Berry K., Baas S., Linher K., Weidman J., Khouri H., Craven B.,		
RL	Bowman C., Dodson R., Gwinn M., Nelson W., Debroy R., Kolonay J.,		
RA	McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;		
RT	"Genome sequences of Chlamydia trachomatis Mohn and Chlamydia		
RN	pneumoniae AR39.";		
RL	Nucleic Acids Res. 28:1397-1406(2000).		
DR	EMBL; AEO01628; AADI8591.1; -		
DR	EMBL; AJ133034; CAB37069.1; -		
DR	EMBL; AP002546; BAA98655.1; -		
DR	EMBL; AE002192; AAF38163.1; -		
DR	TIGR; CP0306; -		
DR	InterPro; IPR003357; -		
DR	InterPro; IPR003368; -		
DR	Pfam; PF02385; OMP; 1.		
DR	Pfam; PF02415; DUF145; 1.		
DR	Signal.		

FT	SIGNAL	1	27	POTENTIAL
CHAIN		28	928	OUTER MEMBRANE PROTEIN.
SEQUENCE		928 AA:	98332 MW:	58910A8F04F12219 CRC64:
Query Match		32.0%;	Score 1376.5;	DB 2; Length 928;
Best Local Similarity		36.6%;	Pred. No. 3.9e-73;	
Matches 349;		Conservative 125;	Mismatches 341;	Indels 139; Gaps 23
OY	1	MKIPLRELLI--SLVPTLSMNLGGAATTE-ELASASNSFDGTTSTTSFSSKTSASADGTIN	57	
DB	1	MKSSILHWFELISSLLPLSLNFSAFVAAYEINLGPINSSCGCTYPPAQQTFA--DGI	58	
OY	58	YFKFSQSVYLENPKNKEHQSTSCFKNDAAGDLNPLGGFSEFTFSNUIDPTTSGAIGSE	117	
DB	59	YVLGDVSTIN-AGSTALTATLASCFFE--TGTNLSPGHGHOPLQINIDA--GACTFTNT	113	
OY	118	AANKVTYTLSEFALSPFLKSPASTVNTGLAIWVKNLSILDNDKVLIDQNF-----	169	
DB	114	AANKLLSFGFVSLDIQF--TNAITGTGAIKSTGACSIQSNVSCYFGQNFSDNGALQ	171	
OY	170	-----	169	
DB	172	GSSISLISLNPILTPAKNKATOKGALYSTGCIITINNLTLSASFSSENTAANGAIYTEAS	231	
OY	170	-----TGQGAINCAGS-----LKIANRKSLSFGNSSSTFGCAI	204	
DB	232	SPTSSKKAISPTINNSTVATSATGAIYGCSTSPKPVVLISDNGELNFTGNRAITSGCAI	291	
OY	205	HKRNTLLSSGGTELFQGNTPAAGKGAIALADSGTSLISGSDGIIIEGNTI--GAT	261	
DB	292	YTDNLVLSGGPFLFKNNASAIPTAALGAIADSGSLSLAGDIFIEGNTYVKAGS	351	
OY	262	G--TVSHASIDIG--TSAKTALRAAGHITTYDDPTTVGSMVADALINSPDGDNKE	318	
DB	352	SSQFTTRNSINIGNTAKVIVOLRASGNTITFYDDPTTISITLALSALMLNPGDLAIGNA	411	
OY	319	YTGTVFSGEKLTEAEKDEKNKTSKLIQNVAFKNGTIVLKGDVVLDSANGFSODANSKLI	378	
DB	412	YQGTIFYSGEKLSEAAEAADNLKSTIQOPLPLLAGQLSLKSGVTLVAKFSQSPSTLL	471	
OY	379	MDLGSILVANTESIELTNIELINIDSLRNKKIKLSAATQOKRIDRPVYVALISDSFQ	438	
DB	472	MDAGTTL-ETAGGITNNVLVNDSLKEKKKATLKQYASQVYTGSLSLVDPSGNAYE	530	
OY	439	NGFLNEDHSYDGI-LELDAGKDVI---SADRSIAVQSPYCGQKATINW---STD	490	
DB	531	DVSWNNPVQFSCLTITLADDPANIHITDLADPLEKRPPIH--WCYQGNMALMSQEDRATMS	588	
OY	491	KRAYSWAKOSFNPTAEQAPLVPNLWGSFIDVRPQNFIELTEGADYEKKRFVWAGIS	550	
DB	589	KATLTLTWTKTGYNPNERRGTLVANTLWLSFYDVRSIQDLVATKVRQSOETRGIMEGIS	648	
OY	551	NVLHSGRENORKFRVSGGAVVAGSTRRMPGGDTLSLGAQOLFAPRKQDVFMMNPFKTA	610	
DB	649	NFPHDQSTIKNGFRISGIYVGAATTLIASDNLITIAACQLGKQRDHFINKNRASAYA	708	
OY	611	GSILRHQDASLVSYSILLBEGGLREILLPYYSKTLPCSFYQOLSXGHDHDKTESLPP	670	
DB	709	ASLHLQHLATL--SSPELRL-----YLPGSSEQPVLFQADISVIYSKNNMKTYTQA	759	
OY	671	PPPTLSTDHTSMGCVYVWAGELGTRAVAVENTSGRGFPRETTTPVKQOAVYSRODSFVELG	729	
DB	760	P-----KGESSWYNDQCALELASSLPHTALSHGELFHAFFPIKVEASYIHDSFEERT	814	
OY	730	AISRPDSDEHLNLAIPLEIKLEK--RFAEQYYHVAAWSPVCSNPNCTTLLSNQGS	787	
DB	815	TLVRSRDSDDLNVSPIGITIFERSRNRASRYEATVIYVADYVRKNPCTTALLINNTS	874	
OY	788	WTKGSNTARQNGIVQASGRSLGAAAELEFGNFGFEMRGSSSYNDAGSKIKF	841	
DB	875	WTKGTGINSRQAGIGRAGIFVAFSPNLLEYTSLNLSMEIRGSSSRSYNDLGGKFOF	928	



RESULT 12  
 09RB71 PRELIMINARY: PRT: 772 AA.  
 AC 09RB71;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
 DE PMP.3.2.  
 GN Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Uuchi K.,  
 Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CMI029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL: AP002545; BAA98226.1; -;  
 DR InterPro: IPR003357; -;  
 DR InterPro: IPR003368; -;  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.  
 SQ SEQUENCE 772 AA; 82931 MW; 484FC56D635801EB CRC64;

Query Match 31.7%; Score 1364.5; DB 2; Length 772;  
 Best Local Similarity 39.7%; Pred. No. 1.5e-72;  
 Matches 334; Conservative 119; Mismatches 284; Indels 105; Gaps 23;

QY 23 GAATTELSASNSFDGTSTSTSSKTSATDGTNYFKSVIENPKTGEQSTCEK 82  
 DB 13 GATATATLSTLT---GTTMSALFSENTSS-----KKGAIQTSALT 50  
 QY 83 NDAAGCDLNFAGGFFSTFNIDATASGAIGSEANKTVTLGSFSLKSPASTVT 142  
 DB 51 ITGNQGVSV-----SDMTSSDSGAITTEA---SVTISNNAKVSFI----- 89  
 QY 143 NGIGAINVKNLSLDNDKYLIDNFTSGD--GGAINCA-----GSLKIANKSLSFIG 194  
 DB 90 -----DNKVTGASSSTTGDMSGAI-CAKVTSTDEVTTLTGQMLFFSN 132  
 QY 195 NSSSTRGAIHTKNTLSSGGEFLF-----QGMPTAAKGAIAADSGTISISDGS 249  
 DB 133 NSTTGAGATYKLLASGGLTLFERNVNGGTAP---KGAIALEDGSLSDSG 188  
 QY 250 DIFEGNTIGAT--GTVSHSAIDLGTSAKITATLRAAGHTIYFDPITVSTSVADALN 307  
 DB 189 DIVFLGNTVYSTPGT--NRSSIDLGTSAKMTALRSAAGRAIYFYDPTTSSSTVTDVLK 247  
 QY 308 INSPDTGDKKEYTGTIVFSEKLTLEADKRNKTSKLLQNAVFKNGTVYLKGDVULSAN 367  
 DB 248 VNTPTDSALQYTGNIIFTEGKLTETEAADSKNLTSLQPLVTLSSGGLSLKGVLTQ 307  
 QY 368 GFESODANSKILMDLSTL--VANTSEITLNTLEIDSLNRNGKIKLSAATAODIRDRP 426  
 DB 308 APTQOQADSRLEMDVGTTLLEADTSTI--ANLVINISSIDAKAKAKIETKATSKNLTISGT 365  
 QY 427 VVLAISDESYONGFLNEDHSYDGLIELDAGKDIVISADRS--INAVOSFYGYGKRW-TI 484  
 DB 366 ITLLDPTGTFYENHSLRNQSYD--ILEKASGVTVSTAVTPDIMEKRFHYGQGMGPI 424  
 QY 485 NMSTD--DKAKVSWAKOSFNPAEQEAPLYPNLLWGSFIDVRPFQNFTELGTGAPYEKR 543  
 DB 425 VMTGASTATATFMTKTYGVIIPNERIGSLVPSNLMAFIDISSIHYIMETANGLGDRA 484  
 QY 544 FVWAGISNVLHBSGRENOKRFHVSAGVAGVASTRMPGGGDTLSLGAQOLFARKOVFMNT 603  
 DB 485 FWCAGISNFFHKDSTYTRGFRHLSSGYVIGNLHTCSDKILISAACQLFGRDRDYFAK 544

QY 604 NFAKTAGSLRLOHDASLVSVSILLGEGRLREILLPYVSKTLPSCSFYGLSYGHTDHRM 663  
 DB 545 NQGVYGGTLYQHNNEYISLPC-----KLRCOSLSYVTELPVLSGMLSTHNDL 598  
 QY 664 KTESLPPPEPLTSTDHSTWCGYVWAGELGTR--VAVENTSGRGFEREYPPVAVQAVYSR 721  
 DB 599 KTVY-----TTPYTVKSGMNDSEFALFEGGRAPICLDESA---LEQYMPFMKLOFYAH 650  
 QY 722 ODSFVELGALSROFSDSHLYNLAIPGIKLEKR--FAQYHYHVAWSPVCRSPKCTT 779  
 DB 651 QESFKEGGTAREFGSSRLVNLALPIGRDKESDCQDATYNNLTGLTVLVNSNPCTT 710  
 QY 780 TLLSNQSMRTKSGNLARAGIVQVQASGFRSLGAAELFGNFGFENGRSSRYVDAAGSKI 839  
 DB 711 TLRISGSMKTFPGNLAARQALVLRAGNHPCFNSEAFESQFSELROSSNRYVNDGAKY 770  
 QY 840 KF 841  
 DB 771 QF 772

RESULT 13  
 086163 PRELIMINARY: PRT: 914 AA.  
 AC 086163;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
 DE OUTER MEMBRANE PROTEIN 5 PRECURSOR (FRAGMENT).  
 GN OMP5.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CD/CWL-029/YR-1310;  
 RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;  
 RT "Identification of two novel genes encoding outer membrane complex  
 associated surface layer proteins in Chlamydia pneumoniae."  
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ001311; CA04671.1; -;  
 DR InterPro: IPR003357; -;  
 DR InterPro: IPR003368; -;  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.  
 KW Signal.  
 FT SIGNAL 1 15 POTENTIAL.  
 FT CHAIN 16 >914 OUTER MEMBRANE PROTEIN 5.  
 FT NON\_TER 914  
 SQ SEQUENCE 914 AA; 95603 MW; 8BF33BAB680F5E3 CRC64;

Query Match 31.5%; Score 1353.5; DB 2; Length 914;  
 Best Local Similarity 33.8%; Pred. No. 8.6e-72;  
 Matches 321; Conservative 160; Mismatches 311; Indels 157; Gaps 23;

QY 1 MKIPLEFLI--SLVPTLSMNLGAAATTELSASNSFDGTSTSTSSKTSATDGTNYV 59  
 DB 1 MKQFSMLVYSTLACFTSCSYFAATAENIGPSDFSGSTNGTYTPK--NTTIGIDY 58  
 QY 59 LTFDITLQNL---GDSALTLKGCFSPTTES--LSFAGKGYSLFLNI--KSSAGAL--SV 111  
 QY 60 FKDSVIENVPKGTGQSTIS--CFKNDAAAGDINFLGGEFTFSSNIDATTASGAAGSE 117  
 DB 59 LTFDITLQNL---GDSALTLKGCFSPTTES--LSFAGKGYSLFLNI--KSSAGAL--SV 111  
 QY 118 AANKVTYLSGFSALSLKSPASTVT-----NGL----- 145  
 DB 112 TTDKNLSLTGFSSTLFLAASVITTPSGKAVCGGDLTFDNNGTILRKQDYCEENGGA 171  
 QY 146 -----GA 147  
 DB 172 ISTKNLSLKNSTGISISFEGNKSATGKGAICATGIVDTITNTTAPTLPSNNIAEAGGA 231



```

RESULT 15
ID 09J542 PRELIMINARY: PRT: 936 AA.
AC 09J542:
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE
DE PROTEIN G FAMILY).
GN PMP.7 OR CP0308.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI-TaxId=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-J138.
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CMI029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39.
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,
RA McClellan G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AP002546; BAA98653.1; -
DR EMBL: AF002193; AAF38165.1; -
DR TIGR: CP0308; -
DR InterPro: IPR002637; -
DR InterPro: IPR003357; -
DR InterPro: IPR003368; -
DR Pfam: PF02385; OMP; 1.
DR Pfam: PF02415; DUF145; 1.
DR ProDom: PD004952; -; 1.
SQ SEQUENCE 936 AA: 100105 MW: 3981DB3C950AF95A CRC64;

```

Query Match 30.9%; Score 1328.5; DB 2; Length 936;  
 Best Local Similarity 32.5%; Pred. No. 2.7e-70;  
 Matches 309; Conservative 158; Mismatches 360; Indels 123; Gaps 14;

```

OY 1 MKIPLRLIISLVPTLSMSNLGAATTEELASNSFSGTSTTSFSSKTSSATDGTNYVF 60
DB 1 MKSSVSWLFFSSIFLPSLSIYAAYTLDDSSNNSTIDSGNGTTFYFSTTDAAGTITSL 59
OY 61 KDSVVIENVPKGTGETSTSCFKNDAAGDLNPLGGSFSTFSNIDATTAAGAAIGSEAN 120
DB 60 LSDVSFQNGALGIPLAGSGELE--AGDGLTFQNGHAKLFAFINAGSAGTAAVSTSAAD 117
OY 121 KYVTLSGFSALSLKPKASTVT--NGIGAINVKNLSLLDNKVLIDNFTSGDGGAIN-- 177
DB 118 KNLLFNDFSRLIISCPDLLSPTGQCALKSVGNLSLTGNSQIITFQNFSSDNGVINTK 177
OY 178 -----CAGSLKIAN----- 186
DB 178 NPLISSTSQPASTSRNQAFTGKGGVYVATGTTIENSPEIVSFSONLAKSGCALYSTD 237
OY 187 -----NKSLSFIGNSSSTRGAIH 205
DB 238 NCSITDNFOVIFDGNAGWEAAQAGCAICCTTDTKVTTLGNKNLSFTNNNTALTGYGALS 297
OY 206 TKNLITSSSGETLFGQNTAPTAAGK--GGAIAIADSGTSLISGDSGDIIFEGNTIGATGT 263
DB 298 GKLVSTISAGGPTLFOSNISGSGAGGAINIAGELALATSLATSGDITFNNQVINGST 357

```

```

OY 264 VSHSIDIGTSAKITLRAAOGHTIYEDPIFYVGSTSVADALINSPTDGNKEYGTI 323
DB 358 STRNAINIIDTAKVTGSRATGOSIYFYDPITNPGTAASDTLNLNLADANSIELEGAI 417
OY 324 VFSGEKILTEAEANDKRNFRSKLLQNAFKNGIVYLGVDVLSANGFSQDANSKLINDLGT 383
DB 418 VFSGEKILTEAEKAIANVSTIQPVLAVLADGLVLRDGYVTFEKRDLQSGFSRIINDGDT 477
OY 384 SLVANNESIELTLEININDSLRNGKKIKLSAATAODIRDRVVALISDESRYONGFLN 443
DB 478 TLSAKENANSLNGLAVNLSDGTNKAALTEADKNISLSTIALIDEGSTYEHNNLK 537
OY 444 EDHSYDGIIEI--DAGKDIYISADSRISINAVQSP---YGYGKNTIM--STDDKATVSW 497
DB 538 SASITYP-LLELTAGANGTITLALSTLQEBEPETHYGYGQNMQLSWANATSSKISIMW 596
OY 498 AKQSENPFAQEAAPLVPNLIMGSFIDVRPPQNTIELGTGAPYEKKRPVAGISNVLHRSG 557
DB 597 TRGYIIPSPERKSNLPLNSLMGNFIDIRSIINQLETKSSGEPEREIMLSGIANPFYRDS 656
OY 558 RENQKFRHVSAGAVGASTRMGCDTSLGFAOLFARQDYFMNTNFAKTYAGSLRLQH 617
DB 657 MPTRHGFRHISGGYALGITAATPAEDQLTFAFCQLFARDRNHITGKNHGDITGASLFTFH 716
OY 618 DASLYSVSITLGEGLREILLPYVSKTLPCSFYCOLSYGHTDHRMKTESLPPPPPTLST 677
DB 717 TELPLFIANTLWKAIRAPVYLSEIQIIPLSFDAKFSYIHTDNHK-----TYT 767
OY 678 DHT---SWGYYWAGELGTRVAVENTSGRGEFFREYTPYKVOAVYSRODSFVELCAISR 733
DB 768 DNSIIQSWNDNAFCADLGLSLPEV--ISVPYLLKEVEPEPKVQYIYAHQODFEYERAEGR 826
OY 734 DFDSDSHLYNATPLTKLKE--RFAQOYHVAMYPDVCRSPKCTTTLISNQSMTKR 791
DB 827 AFNKSSELINVEIPYGVFERDSKSEKGYIDLTLMTLDAVRRNPKCOTSLIADANMAAT 886
OY 792 GSNLARQAGIVQASGFRSGAAAELEFGNFGFEWGRSSRSYNDAGSKIRF 841
DB 887 GTNLARQGSVVRANHPVOVPHMEIRIGQFAFEVYRSSRNNTNLGSKPCF 936

```

Search completed: October 2, 2001, 03:31:23  
 Job time: 263 sec

```
QY 2039 AGATTCCTTAGTGTGATTTTGTGAGCTTTGGAAGATAGAGACTATTGTAGCTA 2098
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2042 ATCTTATCATCTGACAGCTTCTGCCAATATTTCGGGAAGATAGAGATCCTTATAATA 2101
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2099 AGAATCAGGTACAGCTTAGAGAGACTCTATATTACCAGCAACGAACTATATCT 2158
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2102 AAAATAGAGCTTCTGCTATGCGCTCTCCATCTCCAGCATCTAGGACCTTGTCTT 2161
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2159 GTTTCCTTGCAACCTAGGCGCTTGTGCTGTCTTATGTCTACAGATTCCTGCTC 2218
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2162 CTCACAGCTTGTAC-----GCTACCTTCTGTGATCTGAAGTGGACGCTGTCC 2212
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2219 TCTTTTCAGGAACCTTAGCTACACCATACGATACGATCTGAACCAAGTATACAA 2278
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2213 TCTTGATGCTCAGATACAGCTATATCTATAGTAAAAATACATGAACCTATACACCC 2272
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2279 CATATCTTACTGTAAAGAGAGCTGGGGATAGATAGTTCCCTTTAGATTCG--GTG 2335
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2273 AAGCACCAAGGAGAGAGAGCTGCTGTATATATAGAGGTTGCGCTGTGAACCTTGGAGCT 2332
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2336 GAAGAGCTCCGATTTGCTTAGATGAAGTGTCTATTTGAGCAGTACATGCCCTTCATGA 2395
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2333 CCTACACACACTGCTTAAAGCCTAGAGGCTCTCTCCACGCGTATTTCTTTCATCA 2392
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2396 AATTCAGTTTGTCTATGCAATCAAGGAAGTTTAAAGAAC--AGGGAACAGAGCTC 2452
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2393 AAGTAGAACCTTCGTACATACACCAAGATAGCTCAAAAGACGTAACTACTTGCTGATC 2452
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2453 GTGAATTTGGAGTAGCCCTCTTGGAATCTTGCCCTTACCTATCGGATCCGATTTGATA 2512
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2453 GATCTTTCATAGCGGTGATTTAATTAACGTCTGTGCTTATGGAATTAACCTTCGAGA 2512
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2513 AGGAATCAGACTGCAAGATGCAAGTACATCTTACTGTTATTAAGTGTGATCTTG 2572
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2513 GATTCGAGAAACGAGCGCTGCTTAGAAGCTACTGATCTACGTTGCCGATGCT 2572
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2573 TTGCTAGTAACCCGACTGTACGACAACTGCGAATTAGCGGTGATTCCTTGAAAACT 2632
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2573 ATCGTAAGAAATCTGACTGACGACAGCTCTCTTATCAACAATACCTGTGAAAACTA 2632
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2633 TGGGTACGAAATTTGGACAGACAGCTTTAGTCTCTGCTGACAGGAACCAATTTGCTTA 2692
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2633 CAGGAACGATCTCTCAAGACAGCTGTATCGAAGAGCAGGAGATCTTTATGCCCTCT 2692
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2693 ACTCAAAATTTTGAAGCCTTAGCCAAATTTTCTTTGAATTGCGTCAATCTCGCAATT 2752
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2693 CTCCAAATCTTAGGTCAACAAGTACATCTATGGAATTTGATGATCTTACGCGAGCT 2752
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2753 ACAATGTAGACTTAGAGCAAAATACCAATTCTAA 2787
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2753 ACAATGCAAGATCTGAGGTAGTTCAGTCTAA 2787
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: October 2, 2001, 01:27:11  
Job time: 45904 sec

unclassified.  
REFERENCE 1 (bases 1 to 2787)  
AUTHORS Madsen, A. and Birke Lund, S.  
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE  
JOURNAL Patent: WO 9858953-A 13 30-DEC-1998;  
MADSEN ANNA SOFIE (DK); BIRKE LUND SEVEND (DK)  
FEATURES  
source 1..2787  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 815 a 689 c 535 g 748 t  
ORIGIN  
Query Match 18.68; Score 517; DB 9; Length 2787;  
Best Local Similarity 53.18; Pred. No. 8.5e-117;  
Matches 1420; Conservative 0; Mismatches 1180; Indels 75; Gaps 12;  
QY 143 AGATGCTGGAAGTACTTACTTATTTAAGGAAATGTCACCTGAGAAATATTCCTGGAA 202  
DB 158 ATGCAATGAGTACTATATTAATCTAACAGGAGATGCTC---ATACCAATCAGAGAT 214  
QY 203 CAGGACAGCAATCACAAGAGCTGTTTAAACAACACTAAGGCGATTTGACTTTCACAG 262  
DB 215 CTCGACAGCTCTAACCGCTTCTGCTTTTAAAGAACTACTGGGAATCTTCTTCCAAAG 274  
QY 263 GTACGGAAGTCTCTATTTGTTCCAAACGGTGTGATGACAGGAGCTGTAGCAGGGGCTG 322  
DB 275 GCCACGGCTCAATTTCTCTCAAAATATCGATGGGG-----AGCAACTGTACT 328  
QY 323 TTAACACAGCAGTGTGATATTAATCTACCAAGTTATAGGTTTCTTCGCTATCTTTA 382  
DB 329 TTACCAATACAGCTGCAAAATAGCTTCTCTCTTTTCAAGATTTCTATTTGACATA 388  
QY 383 TTGGGCTCTCTGGAAGTGTCAATTAACCGGCAAGAGCCGTAGCTGTACAGGGA 442  
DB 389 TAC-----AAACCAAGATCTACACAGGACAGAGCCATCAAGTCCACAGAG 439  
QY 443 GCTTGAAGTTGACAAATATGTAGTTGCTCTTCAACAAATTTTCAACGATATATG 502  
DB 440 CTGTCTTATCTAGTCACTATATAGTGTACTTGGCCAAATCTTTCTATATACATG 499  
QY 503 GCGGTCTATCAGCGCAAAATCTTTCAATTAACAGGAGTACAAATGCTGTTTT 562  
DB 500 GAGGCGCCCTCAAGAGAGCTATAGTGTATGCTAATGCCAATACCCCACTTAAC---GTTTG 556  
QY 563 CTGAAATACCTCTCTCAAGAAAGGCGGAGCCATTCAGACTCCGATCCCTTAACATTA 622  
DB 557 CCAAAAACAAAGAACCAAAAAGGGGGTCCCTCTATTCACGAGGAGGATTTACAATTA 616  
QY 623 CTGGAACCAAGGGAAGTCTTTTCTGACAACTTCTTCTGAGATTCGAGAGCTCAA 682  
DB 617 ACAATACGTTAACTCAGATCATTTCTGAAATATCCGCGGCAAAATGCGGAGCA 676  
QY 683 TTTTTCAGAGAGCTCGGTACTATTTCTAATATGCTAAAGTTTCTTTATTGACATA 742  
DB 677 TTTTACAGGAGAGTATGATTTTATTAGCAGCAAAAGCAATTAAGTTTATTAAACATA 736  
QY 743 AGGTACAGAGAGGAGCTCTCAACAAGGGGGATATGTCAGAGGTCATCTGTGCTT 802  
DB 737 GTGTGACCG-----CAACCTCAGCTACAGGGGGCCATTACTGTAGTA 781  
QY 803 ATTAACCTAGTACAGATCTAAGGCTCCTCAGCTGAGAAATTCAGATTTACTCTTACGA 862  
DB 782 GTATCATCAGCCCAACACGCTTAACTATACAGCAACGGGGAATCAACTTTATAG 841  
QY 863 ACATATCATGACAACAAGGGAGGAGCTATCTATGTAAGAAAGCTGAGCTGCTCCG 922  
DB 842 GAATATCAGCAATTTACTAGTGTGGGCGATTTATCTAGCAAAATCTAGTCTTCTG 901  
QY 923 GAGACTTACCTATTCAGTAAATAGTGTCAATGAGAGTACAGCTCTTAAGAGTGAG 982  
DB 902 GAGAGCTACGCTTTTAAACAACTCTGCTATATGATCTGACAGCTCCCTTAGAGGAG 961

QY 983 CCATAGCTATCGAAGATAGTGGGAATTTGACTTATCCGCGATAGTGTGACATGTCT 1042  
DB 962 CAATTTGATTTGCTGACTCTGATCTTGTGACTCTTGTGGCTTGTGGAGACATCACTT 1021  
QY 1043 TTTTAGGAAATACAGT-----CACTTCTACTACTCTCGGAGCAAAATTAAGTAGTA 1093  
DB 1022 TTGAAGGAACACAGTATGTCAAAGAGAGCTTCTTGATCTAGACACATCAAGAAATCTA 1081  
QY 1094 TCGACTTAGGA---ACAGATGCAAAATGACAGCTTTGCTGTGCTGTGTAGACCA 1150  
DB 1082 TTAACATCGGAACACCAATGCTAAGATTGACAGCTGCGAGCTCTCAAGGCAATCTA 1141  
QY 1151 TCTACTCTATGATCCGATCACTAATCTACAGATCTTCCACAGATAGATGTCTAAAG 1210  
DB 1142 TCTACTCTATGATCTTATTAACATCACTATCTGACGCTCTTCTGATGCTTAACT 1201  
QY 1211 TTAATGAGACTCCGCGAGATTTCTGACTACAAATTAACAGGGAACATCATCTTACAGAG 1270  
DB 1202 TAAATGCTCGTACCTTGCAGAGGAATCTGATATCAAGGAACCATGATTTCTGAG 1261  
QY 1271 AAAAGTTATCAGAGACAGAGCGCAGATTTCAAAATCTTACTTCAAGCTTACAGC 1330  
DB 1262 AGAAGCTCTGGAAGCAGAAAGCTGATTAATCTCAATTCACAAATTCAGCAAC 1321  
QY 1331 CTGTACTCTTTCAGAGGATCTATCTTTAAACATGGAGTGTGCTGACAGCTCAGG 1390  
DB 1322 CTCTACTCTTTCGAGGAGGCACTCTCTTTAATAGAGGATGCTGTAGTCTAAGT 1381  
QY 1391 CATCTACTCAACAGGAGATTTCTGCTCGAAATGAGAGTAACTACTTCAACCTG 1450  
DB 1382 CTTTGTGCAATCTCGGAGGCTACCTCTCATGATGAGAGGAGCAATTAAGAACCG 1441  
QY 1451 CTGATCTAGCAGCATTAACATTTGCTATTAATCATCAGTTCTATAGAGGTCAAGA 1510  
DB 1442 CTGATGGATCACTATCAAAATCTTGTTCATATGATTTCTTAAAGAGACAGA 1501  
QY 1511 AGGCAAAATAGAAACCAAGCTACGTCMAAAATCTGACTTTTGTGAACATCACT 1570  
DB 1502 AGGCTACGCTTAAGAACACACAGCAAGTACAGACGTCATTTCTGTGAGTCTCTC 1561  
QY 1571 TATTTGACCCGAGCGGAGGAGTATTAAGAAATCAATTAAGAAATCTCAGTCTAGC 1630  
DB 1562 TTGTAGATCTTCTGAAATATGTCACGAAGTGTCTTGTGAATTAACCTCAAGTCTTT 1621  
QY 1631 ACATCTAGAGCTCAAGCTT-----CTGAGCTGTAAAGAGCAGGCTGACTCAG 1684  
DB 1622 CTGTCTCACTTATCTGCTGACAGACCCCGAATATTCACTACAGACTTACCTGCTG 1681  
QY 1685 ATCTTATATGGGTGAGAAATTCATTAAGGCTATCAGGGAATTTGGGCGCAATGTTT 1744  
DB 1682 ATCCCTTAGAAAAAATCTCATCTATTGGGATTCAGAAAGGAAATTTGGCATTTCTGGC 1741  
QY 1745 GGGGAGACGGGGC-----TTCTAGCACTCAACCTTCAACTGAGCTTAATGCTGATATA 1798  
DB 1742 AAGAGATCTCGGAGCTAATTAATCCAAAGCAGGACTTTTACTGAGCAAAACAGGATAGA 1801  
QY 1799 TTCTTAATCCGAGCGCTATGCTCTTATGCTTATCTTACTTATGAAATGATTTATAG 1858  
DB 1802 ATCCGAATCTGAGCGCTGCGGAACCTTAGTGTCTAACAGCTATAGGGGATCTCTGTTG 1861  
QY 1859 ATATTACTCTCTCCATTTATCTATAGGAGCTGAAACGAAGGTTCTCAGAGAGACCTG 1918  
DB 1862 ATGTGCGCTCATTAACAACACTTGTAGCCCTAAAGTAGCCCAATTCGAAGAACTGCG 1921  
QY 1919 CTTTTGTGCTGCTGATTAATTAATCTTTCATATAGATAGTAAACAAACAGAGCGC 1978  
DB 1922 GCATCTGCTGTGAAGGATCTCGAACTTCTTCCATTAAGTATGACGCAATTAATTAAG 1981  
QY 1979 GGTTCGCCAATTTGAGTGGGCTTATGTCATPAGAGGAAACCTAATATCTTGTTCAGATA 2038  
DB 1982 GTTTTGCACATTAAGTGAAGTATGTTAGAGAGGACTACAACTATAGCTTCTGATATA 2041



[illegible]

PLOYVSYFSGEOTGVYFVYDPRKTFESCTQSPRSLIMKRLRELAFINRSTITVPEDEGPROV  
 SPDKTTFPEGGIGQSVYSLNQRKSELPSPIITICGRVYDDGEIEFEALONSGSGLA  
 ELVISTANNIPIRGQSTHGLVETALRTVNTIKANHLKNNKLTALGEGIRGLTA  
 VTSVYANNPEFGQCTQKLGNSVSSVAOVGEALITFEENQIARMTVDYFVAAG  
 QAREAKKRRPELTLRKSADLSALPEKLTIDCLEKDEPKCMTIYEGDSAGSAQOGBD  
 RFIIDALPIRGKLTINERLAKQIPQNOEGTIAALPGCSIGADNFMLSKYRIIL  
 REMOALDGHGIRLLITLFPYRHMTALJENVCYIAADPIKVSCKNFRTILSEKDE  
 SYLIMQKTESSITLSEKRETRJLJENVCYIAADPIKVSCKNFRTILSEKDE  
 KCTIGPLPYLLAPATCHMGCGRTIYSDBEKREALAQETHKFKITELTKAVYFDTQNO  
 LKEYGIDISSYLIIPQKNEVYIGEDSPSCSYTTEEVNTIKNLKRSKIEIORYG  
 IGEAMNDLQMDTMTMNEORTLLHVSLEKDAVEADHIFTMLGMEVPRPREIESHALSI  
 RNNNDI"  
 complement(10812, .11159)  
 /gene="CPJ0276"  
 complement(10812, .11159)  
 /gene="CPJ0276"  
 /codon\_start=1  
 /transl\_table=11  
 /product="CtJ91 hypothetical protein"  
 /protein\_id="BAAG8486.1"  
 /db\_xref="GI:8978650"  
 /translation="MELKRRKRGGSOVQNKRTASPIKHAHYIHTYQELQKIMARP  
 HDAIDAMNODFDKKQKMSQAIGFRDILLVRYKNSLYALLKQTPDNDLINSLYOVA  
 SHVQITREIQLPFLP"

Query Match	Similarity	18.68;	Score 517.2;	DB 2;	Length 300650;
Best Local	Similarity	52.88;	Pred. No. 1e-116;		
Matches 1406;	Conservative	0;	Mismatches 1198;	Indels 60;	Gaps 1
OY 150	TGGACCTACCTACCTATTTAAGGAAATGTCACCTCGAAAAATTTCTGGACAGCAC	209			
Db 197504	TGGAAACGAACATATGCTTATCAGGAATGCTATAT---	AAACATGCTGGAAAAAGCAC	197560		
OY 210	AGCAATCACAAAAMAGCTGTTTTAACACACTAAGGGCGATTGACTTTCACAGSTAA	CGG 269			
Db 197561	AGCATTTAACAGCGCTGCTGCTTTTACAGAAACACGCGGATCTGACATTTACTGG	AAAGGG 197620			
OY 270	GAACCTCTATTGTTTCCAAAGGGTGAATGCAGGGACGATGTAAGAGGGCTCTGT	TAAACG 329			
Db 197621	ATACCTATTTCATTTACCAACAGGTAGAGCGGGTTCCAAATGACAGACTGGC---	CAAG 197677			
OY 330	CAGCGTGTAGATAAATCTTACCAAGTTTATAGGGTTTTCTCGTATCTTTATTTGG	CTC 389			
Db 197678	CACACCTGCTATTAAGCCCTAACATTTCCACAGATTTTCTAACCTTTCCCTTAT	TGGAGC 197737			
OY 390	TCCTGGAACTTCGATTAACCTACCGCAAGAGACCGTTAGCTGCTCTACGGGTAC	TTGAA 449			
Db 197738	TCCTGGAACTACAGTGTGCTTCAGAAAAAGTACCTTTAA---GTCTTCAGAG	AGCCTTTAA 197794			
OY 430	GTTTGACAAAATATGAGTTGCTTCCTCAGCAAAACTTTTCCAAAGC-----	ATTA 500			
Db 197795	TCCTTACGATATATGAAACGATTTCTTTTACCAAAAGCTCTCCAAATAGCTAA	TATACAA 197854			
OY 501	TGGCGTGTCTATTCACCGCAAAACTCTTTTCATTTAACAGGACATACATGTCAG	CTCTGTT 560			
Db 197855	TGGCGAGCCGATTCACCGCAAAACTCTTTTCATTTTCGGAAATACCTCTCTA	TAAACCTT 197914			
OY 561	TTTCGAAATATCTTCCTCAAGAAAGCGGAGCCATTCAACATTTCCGATGCC	CTTACAT 620			
Db 197915	CACATGATATAGCCCAAAAATATGATGGAGCGCATTTATAGCTCTGCGGCTCA	AGTAT 197974			
OY 621	TACTGGAACCAAGGGGAGTCTCTTTTTCGACAAATCTTTCTCGGATTTCTGAG	CTGC 680			
Db 197975	TTTCAGGAACACCGGCCAGTTAGTCTTTATGATATATTAAGAGAAACTGGG	GGTGGGC 198034			
OY 681	AATTTTACAGAAACCTCGGTAGCTATTTCTAATATGCTAAGGTTTCTTTAT	TGACAA 740			
Db 198035	TCCTGGCTTTTAAACCGACCTCTCGATTACCAAAATATAGCTCCCTTTCTT	CTCTGAAA 198094			
OY 741	TAAAGTACAGAGACGAGCTCTCTAACAAAGGGGATATGTCAGAGAGTCTAT	CTGTGC 800			
Db 198095	CACGTCAACACATGCTCTCAGGCAAG-----GGCGGGGCATTTATTTG	198136			



QY	2226	AGGAAACCTTAGTCACACCCATCGGTTATGCAATTCGAAACCAACAGATGATCAACATATCC	2285
Db	669	TGCCCACTAGCTACAGCTACACTATAAAATGATATGATACCTCGCTATATCTCTATCC	610
QY	2286	TACTGTTAAAGGAAGCTGGGGGAATGATATGTTCCGCTTTAGAAATTCGGTGA--AGAGC	2342
Db	609	TGAAGCTCAAGAGCTCTTGGACATAAATCTGGGGGCTCTAGAGCTCGGAGGATCTCTGGC	550
QY	2343	TCCGATTTGCTTAGATGAAGTGCTCTATTTGAGCAGTACATATCCCTTCATGAAATTTGA	2402
Db	549	TTATATATCTCCCTTAAGAACACCGCTTCTTCACAGGATATTTCCCTCTTAAAGTTTCA	490
QY	2403	GTTTGTCTATGACATCAGGAAGGTTTAAAGAAACAGGAAACAGAACTCTGTAATTTGG	2462
Db	489	GGCAGCTACAGCGCCGCAACAAACTTAAAGAGATGGCGCTGAAGCCCGCTCTTTGA	430
QY	2463	AAGTAGCCGCTTGTGAAATCTTGCCCTTACTACTATCCGGATCCGATTTGATAGGAATCA	2522
Db	429	TGAATGAGACCTATGTAAGATGCTCTATCCCTGTCCGATTCGGTTGAAAAATTCGCA	370
QY	2523	CTCGCAAGATGCAACGATCAATATCACTCTGCTGGTTTAACTGTGATCTGTTGTGATTA	2582
Db	369	AGATGAAAAAATAATTTTCAGATTTTCTAGAGCTTACATTTGGTATGTATCGTAAAA	310
QY	2583	CCCCGACTGTACGACAACACTGCGAATTTAGCGGTGATTTCTTGGAAAACTTGGGTACGA	2642
Db	309	TCCCGCTTCGCTACTTCTTAATGTGATGAGAGCGCTCTTGGAACCTTCGATGTAAAA	250
QY	2643	TTTGGCAAGACAGCTTACTTCTCTGTCGTGAGGAAGCAATTTTGGTTTAACGCAATTT	2702
Db	249	CTTCGCAAGCAAGCTTCTTACGAGGTGGAGAGCACTGTACCTCTCCCTCATGT	190
QY	2703	TGAAGCTTTAGGCCAATTTCTTTTGAATTCGCTGGGTGATCTCCGCAATTCATATAGA	2762
Db	189	AGAAGCTCTGCGGGAAGCTGCTTATGAGCTTCTGTGCTGACGACATCTACATGTAGA	130
QY	2763	CTTAGGAGCAAAATACCAATTTCA	2786
Db	129	TTGTGGCTAAGATCACTCATTTCA	106
RESULT 14			
AP002546			
LOCUS	AP002546	300650 bp	DNA
DEFINITION	Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,		08-DEC-2000
ACCESSION	AP002546	AB033782	AB033783
		AB033784	AB033785
		AB033802	AB033803
		AB033804	AB033805
		AB033806	AB033807
		AB033808	AB033809
		AB033810	AB033811
		AB033812	AB033813
		AB033814	AB033815
		AB036079	AB036080
		AB036081	AB036082
		AB036348	AB036349
		BA000008	
VERSION	AP002546.2	GI:10176692	
KEYWORDS			
ORGANISM	Chlamydomophila pneumoniae J138 (strain:J138)	DNA.	
SOURCE	Chlamydomophila pneumoniae J138		
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.		
AUTHORS	Shirai,M., Hirakawa,H., Kimoto,M., Taduchi,M., Kishi,F., Ouchi,K.		
TITLE	Shiba,T., Ishii,K., Hatori,M., Kihara,S. and Nakazawa,T.		
JOURNAL	Comparison of whole genome sequences of Chlamydia pneumoniae J138		
MEDLINE	from Japan and CWI029 from USA		
REFERENCE	Nucleic Acids Res. 28 (12), 2311-2314 (2000)		
AUTHORS	2 (Phases 1 to 300650)		
TITLE	Shirai,M.		
JOURNAL	Direct Submission		
COMMENT	Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.		
	Mutsunori Shirai, Yamaguchi University School of Medicine,		
	Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi		
	755-8505, Japan (E-mail:shirai@pc.cc.yamaguchi-u.ac.jp,		
	tel:81-836-22-2227, fax:81-836-22-2415)		
	on Sep 15, 2000 this sequence version replaced gi:6172290		
	gi:6172292 gi:6172294 gi:6172296 gi:6172326 gi:6172388 gi:6172330		
	gi:6172332 gi:6172334 gi:6172336 gi:6172338 gi:6172382 gi:6172384		

**FEATURES**

**source**

1..300650  
/organism="Chlamydia pneumoniae J138"  
/strain="J138"  
/db\_xref="taxon:138677"  
/note="synonym:Chlamydia pneumoniae (strain J138)"  
144..935  
/gene="CPj0267"  
144..935  
/gene="CPj0267"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAA98471.1"  
/db\_xref="GI:8978641"

**gene**

CDS  
957..1343  
/gene="CPj0268"  
957..1343  
/gene="CPj0268"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAA98478.1"  
/db\_xref="GI:8978642"

**gene**

CDS  
957..1343  
/gene="CPj0269"  
957..1343  
/gene="CPj0269"  
/codon\_start=1  
/transl\_table=1  
/product="dipeptidase"  
/protein\_id="BAA98479.1"  
/db\_xref="GI:8978643"

**gene**

CDS  
complement(1495..2472)  
/gene="CPj0269"  
complement(1495..2472)  
/gene="CPj0269"  
/codon\_start=1  
/transl\_table=1  
/product="dipeptidase"  
/protein\_id="BAA98479.1"  
/db\_xref="GI:8978643"

**gene**

CDS  
complement(2487..3347)  
/gene="yv1C"  
complement(2487..3347)  
/gene="yv1C"  
/codon\_start=1  
/transl\_table=1  
/product="SuA5 superfamily-related protein"  
/protein\_id="BAA98480.1"  
/db\_xref="GI:8978644"

**gene**

CDS  
3656..4387  
/gene="CPj0271"  
3656..4387  
/gene="CPj0271"  
/codon\_start=1



DB 2551 GATCCAGGCTGACGACCACTCTATGTTCTGGGAGATTCTTGCGACATGTGTACA 2610  
 QY 2641 AATTGGCAAGCAAGCTTACTCCTCTCGACGAGAACCAATTTTGGCTTAACCAAT 2700  
 DB 2611 ACCTTGTGACACCACTCTCTTCTGACGTGCGAATCAATCACTCTTCTCAAC 2670  
 QY 2701 TTGGAAGCTTACCAATTTCTTGTGAATTCGCGTGCATCTCGCAATTAATGTA 2760  
 DB 2671 TTGGAAGTTTACGTCAGTTTGAAGTCGAGTTCGCGAGTCTTCTGTAAGTATGCTATC 2730  
 QY 2761 GACTTAGAGCAAAATACCAATTTCA 2787  
 DB 2731 GATCTTGAGAGAAATTCGATTTTCA 2757

RESULT 13  
 AE002193/c  
 LOCUS Chlamydia pneumoniae AR39, section 26 of 94 of the complete genome. 10026 bp DNA BCF 30-MAY-2000  
 DEFINITION  
 ACCESSION AE002193 AE002161  
 VERSION AE002193.2 GI:8163406  
 KEYWORDS Chlamydia pneumoniae AR39.  
 SOURCE Chlamydia pneumoniae AR39.  
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE 1 (bases 1 to 10026)  
 AUTHORS Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Uitterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwin, M., Nelson, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.  
 TITLE Genome sequences of Chlamydia trachomatis Morn and Chlamydia pneumoniae AR39  
 JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
 MEDLINE 20150255  
 PUBMED 10684935

2 (bases 1 to 10026)  
 AUTHORS Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Uitterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwin, M., Nelson, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
 COMMENT On Jun 1, 2000 this sequence version replaced gi:7189234.  
 FEATURES  
 source location/Qualifiers  
 1..10026  
 /organism="Chlamydia pneumoniae AR39"  
 /strain="AR39"  
 /db\_xref="taxon:115711"  
 /note="synonym: Chlamydia pneumoniae AR39"  
 complement(105..2897)  
 /gene="CP0307"  
 complement(105..2897)  
 /gene="CP0307"  
 /note="similar to GP:4376729; identified by sequence similarity; putative"  
 similarity\_start=1  
 /codon\_start=1  
 /product="polymorphic membrane protein G family"  
 /protein\_id="AAF38165.1"  
 /db\_xref="GI:7189235"  
 /translation="MKIPHLKLLISSTIVPILISIAFYADAGSLSPDSDGAGST  
 FTPKSTADANGTIVLSGNVYINDAGKGTALGCCFTETDLEFTGAGYSEFPTVD  
 AGSNAGAATADKALFTFSTNLSFTAGPTVAAGKSTLSAGALNLNDNGTILF  
 SONYNEANNNGGATTAATKLSISGNTSITFTSNAGKILGATISASAGISGNTGOL  
 VPMNKGEGGALGEFASSSITONSITFSGNTATDAAGKATGATCEKGTGPTPLTI  
 SGNKSLTFAENSVTGGALCAHGLDLSAAPTLPSSNRCCGTAGKAGATADSGS  
 LSLSNQGDITFLGNTLTSTPSTRNATILGSSAKITNLRADGQSGTITRTDPLAS  
 TTGASDVLTINPDNSPLDYSGTIVSGEKLISADEAKRAADFTSILKPIALASGTL

ALKGNVELDVNGFTOTEGSTLLMOPGTRIKADTEAISLTKLVLDLSALEGNKSVSIE  
 AGANTITLTLEPLVYODSSNGNEYESHITNOAFTQGLVFTATATASDIYDALILSPV  
 QTPBPHYOGMEATWADTSTAKSGTMTGNTTQGLVFTATATASDIYDALILSPV  
 TLQDQSOANSITQOGRGLMAGSGTANFPHKSGNNOARRKSGYITGSAEDPSEN  
 IFSVAFCOLFGKDKDLFTVENSHYLAISYLQHAFGLPMPSEFSTITMDLPL  
 ILNADLSYTKNDNDITRTSTPEAGSWMATNSGALTEAGSLTLPEAPFOQYF  
 GLVQAVYSRODNFESGAEARAFDDGLVNCISIPVGLIRLEKISDEKNEEISLAYI  
 GDVYKRNDRSRSLVMSGASWTSLCKNLARQAFILASGSHLTLSPVELSGEAAVEIR  
 GSAHLYVDCGRYSF"  
 complement(2941..5751)  
 /gene="CP0308"  
 complement(2941..5751)  
 /gene="CP0308"  
 /note="similar to GP:4376728; identified by sequence similarity; putative"  
 similarity\_start=1  
 /codon\_start=1  
 /translation="MKSSVSWLFEPSIDPLFSSLSIVAAEVLTDSSNNSYDGSNGTTF  
 VPTTDAAGTTTSLSDVSRONACALITPLASGFLDAGDUTROGNALSKAFIN  
 AGSSAGTAVASTAADKNLFDNFSRLTSCPSLSTPGQALSKVSNLTKGSI  
 IPTONFSDNGGVIWTKNFLSGTQFASFSNNOAFTRKOGVYATGTITENSGI  
 VFSQNLKAGSGALYSTDNCSITDNPVIFDGNAMWAAOAGALICCTTDDKTVLT  
 TGNKLSFTNTMTALYFGATISGLKYSISAGTIFOSMSISAGOGGAGALINISAG  
 ELATASGSDITFNNNOYNGSTSRNAINITDTKVTISRAATGOSITFYDPTNPG  
 TAASTDITNLNADANSEIEGATVSEKISPEKRIANVSTINQPAVLAGDL  
 VLKDSVITFYFDLQSPGSRILMDGTTLSAEVALSINGLANSISDGTAKALKT  
 EAADNLSIGDIALIDEGSYENHNKLSATYLLTLTAGAGTITLALSTLT  
 QEPETHRYOGQWMOULANASISKISINWRTGIPPERKSNLPLMSLMGNFDIR  
 SINOLIERKSEPERELMISGLANFPRDMPTRHGFRLISGGYALGATATPAED  
 OLTFAPCOLFADRNRHITGNKNGDRTYSLFHHREGFLDANFAMGATRAPVYSL  
 ISQITPLSPDAKFSYLTNDNKHRTYTPNSITLFGKSRDARADGALTPVLSVPL  
 LKEPPEYKVOITIAHODFYERYAEGAFNKSILINVEITPGYFEDDSKSEKTYD  
 LTMATIDAYRRNPKCOTSLIASDANMAVGINLARQSVRAAHFOVNPMEITFG  
 FAFEVYSSSRNNTNLSGKCF"  
 complement(6013..9843)  
 /gene="CP0309"  
 complement(6013..9843)  
 /gene="CP0309"  
 /note="similar to GP:4376727; identified by sequence similarity; putative"  
 similarity\_start=1  
 /codon\_start=1  
 /product="polymorphic membrane protein G family"  
 /protein\_id="AAF38166.1"  
 /db\_xref="GI:7189237"  
 /translation="MKYSLPWLITSSALVSFSLHPIAMAANTDSSDNYENGSSGAF  
 TAKESDASGTTVLTSDVSITNVSAITPADSCPTNGGALSFEVGAHSLVLTQIAL  
 THDGAALNNTNALSFGSESLIDISAPATGSGKGAICVYNTGATATFNDASVT  
 LOKNTRSEKDAVASVSTIDIAKTATALLDNTNNGALCSTANTVVOGSGVTF  
 SSNTRATDGGGITYSEKSDTIDANTGVTFPSNTRKKGASSDONLALTQVQLQ  
 ENKTTGSAQANPEGCGGALICCYLATVTRKGLAISQNEKSFSTNTTANGATYA  
 TKCTLDGNTLTLEFDONTATACGGAIVYETEDFSLKGSTGVTFSTNTRAKGALYSK  
 GNSLTGNTNLLFSGSNKATGSPNSNSANOEGGALISLESASVTKGMLKEDENY  
 SLSGNTAVSGAIYATKCALHGNTLTFEDGNTATAGATATYETEDFLNGSTV  
 FSTNTRAKGALHFKGNTSPTNKRKALVSNSATATATTTDDECGGALICNISED  
 IATKULTITENESLFINNTAKRSGGIIYARCVISGSESIIPDQNTAFETSGALYSK  
 NLSTIANGPVSTNNSGKGAIYADSGELSEAIDGDIYFSGNRATGISTPVSII  
 LGAGAKITRKLAAAPGHTIYFYDITMEPASGCTIEELIVNPVVAIVPPQPKNGPI  
 ASVYVPAVPAANPTNGITVFSGKLPDSDASIPATTTILNKLINAGNVYLKAGAT  
 LOVYSVQPODSTVPMADGTLTETTTNTDGSIDLKLNVLDLADCKRMITLIVNS  
 TSGGKITSGDKFHNNKSEFYDNPGLKALNLPETLDSSTGTVLDDFNPPLPSMAA  
 PDYOGGSWTLIPRYKAGAGKYVAEMOKALTYPRPEPRATLVPSILMNAVHTHSIO  
 QEIRATMSDAPSHQIWTGIGNAPHQOKENAFRLISKCIYVGSMTTPQETTRA  
 VAFSOLFVSKSDYVVDKISQVYAGSLCAQSSSYVPLHSSLRHVLISVLEPLPEETP  
 LVLHQVYSYKQGHNNHTKILANNQCKDQWSHSFAVVGSLPLDLYRVLTVSPY  
 VKLQVYSYKQGHNNHTKILANNQCKDQWSHSFAVVGSLPLDLYRVLTVSPY  
 DAYRQPHCLISLTNGTSMSPATNLSQAFEAESGHLKILHGLDCFAAGSCELRS  
 SRSVYANGSTYSP"

BASE COUNT  
ORIGIN

2726 a 2051 c 2367 g 2882 t



```

10056 GCCTGAACCTAAGCGAGGAATTTACTACTAAACGTGGTGCACCTGCATACGAA 9997
1389 GGCATCTCAACAGGAGCATTTCTGCTCGAATAGACGATAGCACTACTAGAAC 1448
9996 AGGCTTACTGACAGCGGGGCTCTCTGTATATATGCGGGGCAACGTTAAAC 9937
1449 ---TGTGATAGTACGACCATTAACAAATTTGGTATTAATCATCATCTATAGAGCGTGC 1505
9936 AAGTACAGAGAGAGCTCTTAAACAGCTTTCATCTGATAGACTCTTTAGCGAGG 9877
1506 AAGAAGCAAAATAGAAACCAAGCTACGTCAAAATTCGACTTTATCTGAACCAT 1565
9876 TAAGAAGTTGTAATGTCGTCTCTGACCAAGTAAATATAGCCCTTAGTGCTCGAT 9817
1566 CACTTATATGAGCCGACGCGGACGTTTATGAAATCATAGTTTAAGAAATCTCAGTC 1625
9816 TCTCTTTTGGATACCAAGGGAATGCTTATGAAATTCAGACTTATGGAATAA 9757
1626 CTACGACATCTTAGAGCTCAAGCTCTGGAAGCTATACAGACGCGGACGCTCAG 1685
9756 CTTCATTTTGGACAGCTCTGCTGCGGTATGCAACACTACAGATGTTCCAGCGT 9697
1686 TCTTATATAGGTGAGAAATTCATTAAGCTATCAGGAACTGGGCGCCAAATGCTTGG 1745
9696 TCTTACAGTAGAAGCTCTACGACACTATGCTATCAAGGTAAGTGGGGAATGCTTGGGT 9637
1746 GG-----GGACAGGGGCTTCTACGACTGCACTGCACTCACTGCACTAAACG 1793
9636 TGTATGATACCGCAGACGCTCAAGACGCTAAGACGACGATTAAGCTTGAACCAATACAG 9577
1794 CTATATTTCTTAATCCGAGCGATATGCGCTCTTATAGTCCCTAATAGCTTATGATGAT 1853
9576 CTACCTCCGAATCTGACGCTCAAGACCTTTAGTTCCTAATAGCTTGGGGAATCTT 9517
1854 TATGATATTAAGTCTCTCATTTATCTTATGAGACTGCAAAAGGTTGAGGAGAA 1913
9516 TTTAGACATCCAGGATTCATGATGATAGAGAAAGTGTGCTTGTATGTTTCA 9457
1914 CCGTGTCTTGGTGTGATGATATCTAATCTTCCATTAAGGATATGACAAACAGC 1973
9456 TCGAGGCTCTGGGCTGCGGAGATGCCCAATTTCTTATGATTAAGATTAAGAAAGGGA 9397
1974 ACGGCGGTTTCCGCAATTTGATGCGGTTATGTCATAGAGAGAAACCTACATCTTGTTC 2033
9396 ACGCAATATCCGATCAATATGATGATATGCTATGCGAGGTGACGCCAACTGTTTC 9337
2034 AGATAGATTTTATGCTGATTTTGTGAGCTCTTGAAGAGATAGAGACTTTGT 2093
9336 TGAACCTTAATAGCTTGTGCTTGTGCACTTGTGATGCGATTAAGATTTCTTGT 9277
2094 AGCTAAGATCAAGGATAGCTAGGAGAACTCTATTTACCAACAGAAACCTTA 2153
9276 CCGTAAATATCACTAGATACCTATGAGAGAGCTTCTATACCAACATTAACAGATG 9217
2154 TATCTCTCTCTGCAACATCAAGCGCTTGTGCTTATGTTCTTACAGAGATTC 2213
9216 TATGAGGTTTATAGTGTCTCTTATGATAAA---CTTCTGCTGCTTGGAGCATTAAC 9160
2214 TGTCTCTTTAGGAAACCTTAGACCATACGCAATACGATTAAGATTAAGAAACAGTA 2273
9159 CCGGTTTGAAGAGGAGCTGCTTATAGCAGCTAGTAAATGATTAAGCAAAAGTA 9100
2274 TACACATATCTACTAGTTTAAAGAGAGCTGGGGAATGATAGTTAGCTTGAATTCG 2333
9099 TACTGCTATCTCGAGGTAAAGGTTCTGGGGATATATGCTTTAATGATATGTTGG 9040
2334 TGAAGAGCTCCGATTTGCTTGAAGAGAGTCTATTTGAGCAGTACATGCTTCA 2393
9039 AGCTTCTCTCATCTATTCCTGAATCTGATGATGTTGTTTACTATGCTACATACAT 8980
2394 GAATTCAGTTTGTCTATGACATCAGAGGTTTAAAGAAAGGGAACGAAAGCTG 2453

```

```

8979 CAAACTGAATCTGACCTATATACGTCAGACAGCTTCGAGAAAGTACAGAAAGAG 8920
2454 TGAATTTGGAAGAGCCGCTTGTGAATCTTGCTTACCTATCGGAGTCGATTTGATA 2513
8919 ATCTTTGATGACAGCAACCTTCTCAATTTATCTTTGCCATTAAGGGGTCAAGTTGAGA 8860
2514 GGAATTCAGCTGCAAGATGACAAAGTACATCTAATCTTGGTTATATCTGATCTGT 2573
8859 GTTCTGTATGTAATGACTTTTCTTATGATCTGACTTATATCTTATGCTGATCTTAT 8800
2574 TGTGATTAACCCGAGCTGATGACAAAGCTGCAATTTACGGTGTATCTTGGAAACCT 2633
8799 CCGCATGATCCCAATGACATACAGACACTTGTAAACAGCGGAGCTCTTGGGAACTTA 8740
2634 CGGTAGATTTGGCAAGCAAGCTTACTCTGCTGTCAGAGGAACATTTTGTCTTAA 2693
8739 TCCATTAATCTTAGCAGCAGAGCCCTTGAAGGCTGCAAGGAGTCACTACGCTTCTC 8680
2694 CTCAATTTTGAAGCCCTTATAGCCATTTTCTTGTGAATTCGCTGGGTCACTGCAATTA 2753
8679 TCCATATGTTGAAGTCTGCGGAGTTTGTCTTGAAGTTCGATGATCTCAAGGATTTA 8620
2754 CAATGATCTTAGAGCAAAATACCAATCTA 2786
8619 TATGTAGATCTTGGGCTGATGATTCATTTCA 8587

```

```

RESULT 12
A81837
LOCUS A81837 2757 bp DNA PAT 21-JAN-2000
ACCESSION Sequence 11 from Patent WO9858953.
A81837
VERSION A81837.1 GI:6731869
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
REFERENCE
1 (bases 1 to 2757)
AUTHORS
Madsen,A. and Birkelund,S.
TITLE
NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL
PATENT: WO 9858953-A 11 30-DEC-1998;
MADSEN ANNA SORIE (DK); BIRKELUND SVEND (DK)
FEATURES
source
Location/Qualifiers
1..2757
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 770 a 564 c 642 g 781 t
ORIGIN

```

```

Query Match 19.0%; Score 528.6; DB 9; Length 2757;
Best Local Similarity 53.5%; Pred. No. 1,2e-119;
Matches 1428; Conservative 0; Mismatches 1149; Indels 90; Gaps 12;
142 GAAGATGCTGAACCTACCTATTTAAGGAATGTCAGCTAGAAATATTCCTGA 201
160 GATGCTAGTGGCAGCACCTATATTCGATGGGAGTGTCCATA---AGCCAAAGAGG 216
202 ACAGGACAGCAATCAACAAAGCTGTTTAAACAACATAAGGCGATTTGACTTAC 261
217 AACAACAGAGCTTAAACCAAGTGTTTTCTAATACGTCAGAGAAATCTTACCTTCT 276
262 GGTAAAGGAGACTCTCTATTTGTTCCAAAGCTGATGACAGGAGCTGACAGGCGTCT 321
277 GGAAGAGGATTTCTCTCATTTTGAACATATTAATTTCTGCTACGTGACAGGTGTT 336
322 GTTAACAGAGGCTGATTAATCTACACGTTTATAGGTTTCTTCTGCTATCTTT 381
337 GTTAGCAATACAGACAGCTTCTGATTAAGAAATCTCGAGATTTTCAACCTCTGGATG 396
382 ATTGCTCTCTGGAAGTTGATTAATCTACGCGCAAGAGAGGCTTACGCTTACGGCT 441
397 CTTCGAGCTCT-----AGGACACAGGTAAGAGGACATTA---AAATTAACGAT 444

```





```

db      2754   TAATGTAGATCTTGGGCGTAAGTTCCATTCTA 2786
||||||| | | | | |||||
RESULT  11
LOCUS   CPN133034/c
DEFINITION
CPN133034    17280 bp          DNA                      BCT       27-FEB-2001
Chlamydia pneumoniae GGAI-protein family cluster A, genes omp12
(partial), omp11, omp10, omp5, omp4, omp13 and omp14 (partial) and
ORF1, strain VR1310.
ACCESSION
AJ133034
VERSION   AJ133034.1 GI:4455881
KEYWORDS
hypothetical protein; omp10 gene; omp11 gene; omp12 gene; omp13
gene; omp14 gene; omp4 gene; omp5 gene; ORF1: outer membrane
protein; outer membrane protein 11; outer membrane protein 12;
outer membrane protein 14; outer membrane protein 4; outer membrane
protein 5.
SOURCE     Chlamydophila pneumoniae.
ORGANISM   Chlamydophila pneumoniae.
REFERENCE  1 (bases 1 to 17280)
AUTHORS   Daugaard,L., Hjerno,K., Knudsen,K., Madsen,A.S., Christensen,G.
and Birkefeld,U.S.
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 17280)
AUTHORS   Boesen,T.
TITLE      Direct Submission
JOURNAL    Submitted (21-JAN-1999) Boesen T., Department of Medical
Microbiology and Immunology, University of Aarhus, The Bartholin
Building, DK-8000 Aarhus, DENMARK
FEATURES
source     location/Qualifiers
           1..17280
            /organism="Chlamydophila pneumoniae"
            /strain="VR1310"
            /db_xref="taxon:83558"
           1..841
            /gene="omp12"
           <1..841
            /note="omp12"
            /gene="putative"
            /codon_start=2
            /transl_table=11
            /product="outer membrane protein 12"
            /protein_id="CAB37067.1"
            /db_xref="GI:4455882"
           /db_xref="SPRREMBL:Q9Z5Q4"
            /translation="EDNIERYRHNSGVALGITATTPAEDQLTFEAFQLFARDNRHLITGK
KNIGDYVIGASLYFHHTGEGLDIFLMGKATRAPMWVSEISOLISFEAKFSYLRHD
NHKKTYVTNSIIRKSGMRNDAPCADIASIPVTSIVPYLIKVEPPEPKVOYIAHODD
FYRRABGRARFNKSELINAEIPIPGVTFERSKSKEGYTDLTMLTLDAIRNRPCOTS
FLISDNMMAYGTNLARKGFVRAAHNPVNPMELFGPAFEVRSSRYNTNLGSK
FCF"
           885..967
            /gene="omp11"
            /gene=".3677"
            /gene="omp11"
            /note="putative"
            /codon_start=1
            /transl_table=11
            /product="outer membrane protein 11"
            /protein_id="CAB37068.1"
            /db_xref="GI:4455883"
           /db_xref="SPRREMBL:Q9Z393"
            /translation="MKTIPLHKLLISSLVLPILLSIATVYGADASLSPDSDGAGST
FEPKSYADNGTVYVSLGNVYINDACKGALTCCEFTETGDLTFMGKGSFENYVD
AGSNACAASAATDKALTFEGFENLSFIAPAGRTVASGKSTLSAGALLDTGNTLIIF
SOVSWEANNGAITTKRLTSLTGNTSLPTFSNSKKIKGAGIYSSAAASISGNTGL
VFNNKGTEGGCALGFEPAASSITONSLEFFSGWTATDDAARGGAITEKETGPPTLT
SGKSLTFEANS SVTGGCAICAHGLDISAAGPILFSNNRGNTAACGKAIALADSGS
LTSANQGDTFLGNLTISTAPTSTRNALIYLGSSAKITNLMRAOGSIYFDPIASN
TTSASVLITINQDNSPLDYSGTIFYFSGEKLSADEKADNFTSIIKOPLASGTL

```

[illegible]



QY	618	CATACGTGGAAACCAACAGGGGAACCTCTTTCTTGACATCTCTCTCTGGAACTCTGGAGC	677
Db	630	TATTACAAATATATACGGCTCTCTACCCCTCTCTCGAACAATATGCTGAAGCTGCAGGGTG	689
QY	678	TGCATTTTATACAGAAACCTCGGGTGACATATTTCTAATAATGTGTAAGTTCTCTTATATGA	737
Db	690	AGCTATAAATATGACAGAAACGTGTCAATATACAGGAATATGTCCTGTATTTTCTGA	749
QY	738	CAATTAAGTCTACAGGACCGAGCTCTCAACACGGGGATATGTACAGAGGCTCATCTG	797
Db	750	AAATAGTGTGAC--ACCGACCGCAGAGAAATGAGAGAGCTCTTTCTTGAGATGCCG----	802
QY	798	TGCTTTATAAACTAGTACATATCTTAAGGTACCCCTCACTGAAATACAGATTTACTCTT	857
Db	803	-----ATGTACAAATCTGGGATATCAGATGTAACCTT	836
QY	858	CAGCAACAATATCATCGACACACGGGGAGAGCTATCTATGTGAAAAAGCTGCACTGGC	917
Db	837	CTCAGGAACCAACACTGTACTACTAATGGCGGACATTTATGTGAAGAAGTTTACACTGGC	896
QY	918	TTCCGAGGAGACT-----TACCCTATTCTAGTGAATATGTCATATGAGGAGTACAGCTCC	971
Db	897	TTCCGGGGGGGGGGGGGGGTATCTCTCTTTCTTCAACAATATAGTCCAGGTACCATCTGCAG	956
QY	972	TAAAGGTGAGACCATAGCTATCGAAGATAGTGGGGAATGTAGATTATCCGCCGATAGTG	1031
Db	957	TAAATGTGAGACCATTTCTTCTTACTGCGACGTGGAGAGTGTACTCTTTTCAGCAGAAAGC	1016
QY	1032	TGCAATTTGTCTTTTATAGGAAATCACTCACTCTTCTACTCC---TGGAGCAATAGAG	1088
Db	1017	GGCAATTAACCTCTCAATGGGATGCCATTTGTGCAACTACACCAAACTACAAAAAGAA	1076
QY	1089	TAGTATTCAGCTTATAGGAACGATCGTCAAGATATGACAGCTTGTGCTGTGGTGAAGC	1148
Db	1077	TTCTATTGACATATAGGATCTACTGCAAAAGATCAACGAAATTTACTGCAATATCTGGGCATAG	1136
QY	1149	CATCTACTCTATGATATCCATTAACATACAGAGATCTTCCACACAGTTACAGATGTCTTAA	1208
Db	1137	CATCTTTTCTACAGATCCGATCTACTCTCAATACGGCTGGGATTTCTACAGATACTTAA	1196
QY	1209	AGTTATGAGATCCGGGACATTTCTGCACTACAAATATACAGGAACATCATCTTACAGS	1268
Db	1197	TCTCATATAGGCTGATGACAGTATATAGTACAGATTTATAGTGGGTGATTTGTTTTCTGG	1256
QY	1269	AGAAAGATTATCAGAGACAGAGGCGCGAGATTTCTAAAAATCTTACTTGAAGCTACTACA	1328
Db	1257	TGAANAACCTCTGTAAGATATAGCAAAATTTGCAGACAACCTCATCTTACGCTGAAGCA	1316
QY	1329	GCCGTATACCTTTTACGAGGAGTCTCTATCTTTAAAAACAATGAGAGTCTGCGAGATCA	1388
Db	1317	GCCGTATCTCTAATCTGACAGAAATTTTAACTTAACTGCTGTGATCTCTCTGATTCAGTA	1376
QY	1389	GGCATTCATCAACAGGACAGATCTCGTCTCGAAAATGACAGTGAAMTACTCTAGAAC	1448
Db	1377	AGGCTTACTCAACACGGGGTCTCTCTGTATTATATGATGGGCGACAAAGTTAAAAAGC	1436
QY	1449	---TGCATATCTACAGACCACTTAACAATTTGGTCAATTAACATCAAGTCTATATAGAGGTGC	1505
Db	1437	AAGTACAGAGAGGAGCTCATTTAAACAGTCTTTCACATTCCTGTAGACCTTTTAAAGCGAGG	1496
QY	1506	AAAGAAAGCAAAATATGAAACCAAAACCTACGTCACAAAATATGCAATTTATCTGGAACAT	1565
Db	1497	TAAAGATGTATATGTCTCTCTCTCTGACACAGTAAAAATGTAGCCCTTAATGGTCCGAT	1556
QY	1566	CACTTTATGAGACCGACGGGCACTTTTATGAAAAATCATAGTTTAAAGAAATCTCTCAATC	1625
Db	1557	TCTTCTTTTGAATTAACCAAGGGAATGCTTATGAAAAATCAGCACTTAGGAAAAAATCTAAGA	1616
QY	1626	CTAGACATCTTAGAGCTAAAGCTCTGGAACGTATACAAAGCAACCGCAGAGTACCTCAGA	1685
Db	1617	CTTTTATTTGTGACGCTCTCTGCTGTGGGTACTGCAACACTACAGATGTTCACAGCGGT	1676

QY	1686	TCCTTAATGGGTGAGAAATTCATTCCATTCAGCCCTACACGGGAACCTGGGGCCCAATGTGTTG	1745
Db	1677	TCCTACAGTACCAACTCCCTACAGCACTATGGGTATCTCAAGTACTTGGGGAATACCTTGGCT	1756
QY	1746	GG-----GGACAGGGGGCTTCTACAGCTGCACCTTCAACTGCATPMAACG	1793
Db	1737	TGATGATACCGCAGACACTCCAAAGACCTAAGACAGGACATTAGCTTGGACCAATACAGG	1796
QY	1794	CTAATATCTTAATCCCGAGCGTATCGGCTCTTATAGTCCATAATAGCTTATGGAATGCATT	1853
Db	1797	CTACCTTCGGAATCCTCGAGGGTCAAGGACCTTATAGTCTTAATAGCTTTGGGGATCTTT	1856
QY	1854	TATAGATTTAGCTCTCCATTATCTTATAGAGACTGCAAGCAAGAGGTTCCAGGAGA	1913
Db	1857	TTCAACACTCCACAGCGATCAAGGTGCATAGAGACAAGTCTTTGACTCTTTGTTCAGA	1916
QY	1914	CCGTGCTTTTGGTGTCTGTGATTTATTACTTCTCCATAAGATAGTACCAAAACAG	1973
Db	1917	TGAGAGCTTCTGGGGCTCGGGAGTCCGCAATTTCTTAGTAAAGATPMAAGAAAGGGGAAA	1976
QY	1974	ACGGGGGTTTCGCCATTTGAGTGGCGGTATGTCTATAGGAGAAACCTACATCTTGTTC	2033
Db	1977	ACGCAAAATACCGTATAACTGTGTGATATGCTATCGSAGGTGACGGCAAACTTGTTC	2036
QY	2034	AGATPAGTTTCTAGTCTGCATTTTGTACGCTCTTTGGAAGAATPAGAACTACTTGT	2093
Db	2037	TGAAAACCTAATATAGCTTTTCCCTTTGCCAACTTTTGGTATCGATPMAAGATTTCTTGT	2096
QY	2094	AGCTPAGAAATACAGTACAGTCTACGAGAGACACTCTATACAGCAACAGCAACCTA	2153
Db	2097	CGCTAAAATATATCTATACCTATACCTATACGAGCGCTTCTATATCCACACATTAACGAATG	2156
QY	2154	TATCTCTCTTCCTTGCAAACTACGCGCTTGTTGTTGTCTTATGTTCTTACAGAGATTC	2213
Db	2157	TAGTGGGTTCATAGTGTGTCTCTTAGATAAA---CTCTCGGCTCTTGGAGTCATPAAAC	2213
QY	2214	TGTTCTCTTTTCAGAAACCTTAGCTACACCCATACGATTAACGATCTGAAAACCACTA	2273
Db	2214	CCTGCTTTTAAAGAGGACGCTCCGCTTATACCCACGCTACGATTAATGATCTGAAACCACTA	2273
QY	2274	TACAACATATCTCTGTCTTAAAGAACGTTGGGGAGTATGATTGCTTACGTTTGAATTCGG	2333
Db	2274	TATGCGCATCTAGGTGAAAGGTTCTTGGGGAATATGCTTTTAAACATGATGTTGG	2333
QY	2334	TGGAAGAGCTTCGATTTGGCTTAGATGAAGTGCCTATTTAGAGCATGACGCCCTTAT	2393
Db	2334	AGCTTCTTCTCATCTTATCTGATACCTGAAATACCGCATGTGTTTGAATACCTPATGCTCATATAT	2393
QY	2394	GAATTCGATTTTCTCTATGCACATACAGAGGTTTAAAGAACAGGACACAGAGCTCG	2453
Db	2394	CAAACTGAAATCTGACATATATACGTAGAGCACCTTCTCGGAAAGAGTACAGAAAGAG	2453
QY	2454	TGAATTTGGAATGTACCGGCTGTGATGATCTGTGCTTACCTACGAGATCCGATTTGATTA	2513
Db	2454	ATCTTTTGAATGACAGCAACTCTTCAATTTATCTTTTGCTTATAGGGGTGAAGTTTGAGA	2513
QY	2514	GAATACAGACTGCCAAGATCAACGTAACATCTTGTGTTATACTGTGATGATGTT	2573
Db	2514	GTTCTCGATTTGTAATGACATTTTCTTATATGATGATGACTTATATCTTATGTTCCGATCTTAT	2573
QY	2574	TCGTATGTAACCCGCACTGTACGCAACACTGCGAATTAACGGTGTATCTTGGAACCTT	2633
Db	2574	CCGGAATGATCCCAAAATGACACTACAGCACTGTAAATACAGAGGCTCTTGGGAACCTTA	2633
QY	2634	CGGTACGAATTTGGCAGACAAACCTTATGTCCTGTGTGAGGGAACCAATTTTGGTTTAA	2693
Db	2634	TGCCAATPACTTACACGACAGAGGCTTGCAAGTGCSTGAGGCACTACATACGCTTGTTC	2693
QY	2694	CTCAAAATTTTGAACCTTTATGCAAAATTTTCTTGAATTCCTGTGGTATCTCCGAATTA	2753
Db	2694	TCTCATGTGTTGAAGTGTCTCGGCCAGTTTGTCTTGAAGTGTCTGATCTCTACAGGATTTA	2753
QY	2754	CAATGTAGACTTATAGAGCAAAATPACCAATTTCTA	2786





Db 206274 TCTCTTTTGGATAACCAAGGAATGCTTATGAAAAATCAGACTTAGAGAAAAATCCACGA 206215  
OY 1626 CTACGACATCTTAGAGAGCTCAAAAGCTTCTGGAACTGTAAACAGACCGGAGCTCCGACA 1685  
Db 206214 CTTTTCATTTTGGACAGCTCTCTGCTCTGGGTACTGCACAACTACAGATGTTCCAGCGGT 206155  
OY 1686 TCCATTAATGGGTGAGAAATTCATTTACGGCTATCAGAGGAAGTGGGCGCAATTGTTGG 1745  
Db 206154 TCCCTACAGTAGCACTCCACGACACTATGGGTATCAGGTACTTGGGGATATACCTTGGGT 206095  
OY 1746 GC-----GACAGGGGCTTACAGACTGCACCTTCAACTGAGCTTAAACTGG 1793  
Db 206094 TATATGATACCGCAAGACCTCCAAAGACTTAAGACAGCACTTACCTTGGACCAATACAG 206035  
OY 1794 CATATTCCTATCCCGAGCGGATAGGGCTTATGCTCCCTATAGCTTATAGTAAATGCAAT 1853  
Db 206034 CTACCTCCGAATCCGACGCGCAAGAGACCTTATAGTCTTAATAGCTTTGGGATCTTT 205975  
OY 1854 TATATATTTACCTCTCTCCATTTATGAGACTGCACAAAGAGGTTGCAGAGGAGA 1913  
Db 205974 TTTACAGACATCCAAAGGATTCAGAGGTCTATAGAGAAAGTGTGGACTTGTTCAGA 205915  
OY 1914 CCGTCTTTTGGTGTGCTGGATTTCTACTTCTTCCATAAGATAGTACAAAAACAGC 1973  
Db 205914 TCGAGGCTTCTGGGCTGCGGGAGTGCCTCAATTTCTAGATAAAGATAAAGAGGGGAAAA 205855  
OY 1974 AGCGGGGTTGCGCAATTTAGAGGGGTTATGTCATAGAGGAAAACTTACATCTTGTTC 2033  
Db 205854 AGCGCAATACGCTCTAAATCTGATGATATCTTATCGAGGTGCAGCGCAACTTGTTC 205795  
OY 2034 AGATAGATCTTCTAGTGTCTGATTTTGTACGCTTTTGGAAAGATAGAGTACTTGT 2093  
Db 205794 TCAAACTAATATAGTGTGCTTTTGCACACTTGTGTAGCAATAAGATTTCTTACT 205735  
OY 2094 ACTTAAGATCAAGTACAGTCTACGGAGGAAGTCTCTATTCCAGCACAAAGAACTTA 2153  
Db 205734 CCGTAAAACTTACTGATATATGACAGAGCTCTTATATCAACACTTACAGAAAG 205675  
OY 2154 TATCTCTCTCTGCAAACTACGGGCTTGTGCTTATGTTTCTTACAGAGATTC 2213  
Db 205674 TATGTGGTTCAATAGTGTGCTCTTAGATAAA--CTTCTGCTCTTGGAGCATTAACC 205618  
OY 2214 TGTCTCTTTTACGAAACCTTAGCTACACCCATCGATTAACGATCTGAAACCAAGTA 2273  
Db 205617 CCTCGTTTAAAGAGGACGCTGCTTATAGCACAGTAAATGATCTGAAGCAAAAGTA 205558  
OY 2274 TACACATATCTCTACTGTTAAAGGAAGCTGGGGGAATGATTTGCTTGAATTCG 2333  
Db 205557 TACTCGTATCTCTGAGGTGAAGGTTCTTGGGGGAATAATGCTTTTAAACATGATGTTGG 205498  
OY 2334 TGAAGAGCTCCGATTTGCTTATGATGAAGTGTCTTATTTGAGCAGTACATGCCCTCAT 2393  
Db 205497 ACCTTCTTCTATCTTCTTATCCGAAATACCTGCATGTTTGTGATACCTATGCTCCATACAT 205438  
OY 2394 GAAATTCAGTTTGTCTATGCAATCAGAGAGGTTTAAAGAACAGGAGACAGAGCTCG 2453  
Db 205437 CAACATGATCTGACCTATATATAGTCAGACAGCTTCTCGAGAAAGATACAGAAAGAG 205378  
OY 2454 TGAATTTGGAATGACCCGCTTGTGAATCTGCTTACTATCAGGAGTCCGATTTGATTA 2513  
Db 205377 ATCTTTTGTGACAGCAACCTTCAATTTATCTTGGCTATAGGGGGTGAATTTGAGAA 205318  
OY 2514 GGAATCAGACTCCAGAGATGCAACGTACATCTTGGTTATCTAGTGCATCTTGT 2573  
Db 205317 GTTCTCTGATTTTATAGATTTTCTTATGATCTGACTTATCTTATGCTTCCGATCTTAT 205258  
OY 2574 TGTATTAACCCCGAGCTGAGCAACACTGCGAAATTAAGCGGTGATCTTGGAAAACTT 2633  
Db 205257 CCGCATATGATCCCAATGACATACAGACTTGTAAACGCGAGGCTCTTGGAAAACTTA 205198  
OY 2634 CGGTACGATTTGGCAAGACAGCTTACTCTGCTGAGGAGAACATTTTGCCTTAA 2693  
Db 205197 TGCATATTACTTAGCAGCAGAGGCTTGCAGAGTGGCTGAGGCGAGTCACTACGCCCTTCTC 205138

OY 2694 CTCAATTTTGAAGCCTTATAGCCAAATTTCTTTGATTTGCGTGGTCATCTCGCAATTA 2753  
Db 205137 TCCATATGTTGAAGTGTCTGGCCAGTTTGTCTTTGAAGTGTCTGATCTCTACGATTTA 205078  
OY 2754 CAATGTAGACTTAGAGCAAAATACCAATCTTA 2786  
Db 205077 TATGTAGATCTTGGGGGTAAGTCCAAATCTTA 205045  
RESULT 9  
AE001628/c DNA BCT 01-DEC-2000  
LOCUS Chlamydia pneumoniae section 44 of 103 of the complete genome.  
DEFINITION  
ACCESSION AE001363  
VERSION AE001628.1 GI:4376730  
KEYWORDS  
SOURCE  
ORGANISM Chlamydia pneumoniae CWL029.  
REFERENCE Chlamydia pneumoniae CWL029  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
AUTHORS 1 (bases 1 to 10757)  
Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,  
Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.  
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
MEDLINE 99206606  
PUBMED 10192388  
REFERENCE 2 (bases 1 to 10757)  
Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,  
Grimwood, J., Davis, R.W. and Stephens, R.S.  
AUTHORS Direct Submission  
TITLE Submitted (01-DEC-1998) Program in Infectious Diseases, University  
JOURNAL of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
FEATURES  
source location/Qualifiers  
1. 10757  
/organism="Chlamydia pneumoniae CWL029"  
/strain="CWL029"  
/db\_xref="taxon:115713"  
106. 2892  
/gene="pmp\_9"  
/note="Cpn0447"  
106. 2892  
/gene="pmp\_9"  
/codon\_start=1  
/transl\_table=11  
/product="Polymorphic Outer Membrane Protein G/I Family"  
/protein\_id="A0D18591.1"  
/db\_xref="GI:4376731"  
/translation="MKSLHMFLLISSSLALPLSLNFSFAAVEINLGPNTSFGPGT  
YTPAQITNADGITYNLGDVSTINAGSPITALTASCFEETGNSLFGHGOFLQNI  
DAGANCPTNTAANKLISFSGESYSLIQTNNATGTGAKISTGACSIOSNYCYFG  
NFSNDGALIGSSISLSLNPILPARKATOKGALSTGISTNNTLNSFSENT  
AANNAGATYTERASSPISNNKALISPTNNSTPNSATGAGLYGSSPAPPYLTIDNGE  
LNFIGNATISGALITDNLVLSGGPILFNNSAIDPAALGALAIADSGSLSLA  
LGDDTFEGANTVVGASSQTTTNSINIGNTAKIVOLRASQGNITFYDITTSIT  
LAALSLALNPLNDLGNAPAYOGTIVFSEKISEAPAAADNLRKSTIQDPLRLAGQLS  
LKSQVTLVAKSFSPGSLTMDAGTILETDGITINNLVNLVNSLKTAKTLAKTO  
ASQVTLVLSGLSDVPNGVNEYDVMMNPQVSCFLTEADPANHIDPLADPLEKN  
PIHNGYOGNMAISMEEDTATSKAATLTWTKGYVNPNERGGTLVANTLKGSPVRS  
IQOLVATKVRSGQETRGIRGSGISNFKDSTKIKGRHISAGIVCATTTTASDL  
ITAAFCQLEGRDRHFIKNKRASAVAAALHLOHATVSSPSLRLRLPDSSEQVPLD  
AOISYISKNTMKYITTOAPRGESSWYDNGALBELASLPHALSHBGLFRAYPEFK  
VEASVYIHDSFKERTTILVRSFDSGLDNLNVSPIGIIPEERSRNERASVEATVIYVAD  
VYRKNPDCTTALLINNTSMKTTGTMLSRQAGIGRAGRIRYARSPNLEVTSNLSMEIRGS  
SRSNADIGGKRFQ"  
complement(3233..4348)  
/gene="yxj6\_2"  
/note="Cpn0448"  
complement(3233..4348)  
/gene="yxj6\_2"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical Protein"







OY	858	CAGCAACAATTCATCAGCAACACCGGGAGAGCTATCTATGTGAAAAAGCTCGAACTGGC	917
OY	858	CTCAGGAACCAACACCTGTACTAATATGGCGGACCAATTTATGTTAAGAAAGTCTACCTGGC	918
OY	918	TTCCGGAGAGACT-----TACCCATTTCCAGTACAAATTAAGTGCATATGGAGGATACAGCTCC	971
Db	5922	TTCCGGGGGGGGGGGGGGGGTATCTTCCTTTCTTTCCTAACATATATAGTCCAAAGGTACCACTGGCAGG	5961
OY	972	TAAAGTGGAGCCCATACCTATCGAAGATAGTGGGAATTAAGTTATTCGCGGATATGGG	1031
Db	5982	TAAATGGTGGACCAATTTCTATACCTGGAGCGAGTGGAGTGTAGCTTTTGACGAGAAGCAGG	6041
OY	1032	TGCATCTGCTTTTATAGGGAATACAGTCACTTCTACTCTCC---TGGAGCAATATAGAG	1088
Db	6042	GGACATTTACCTTCATATGGGAATGCAATGTGTGGCACTACCAACAACACTACGAAAGAGAAA	6101
OY	1089	TAGATTCGACTTTAGGAGACGAGTGGCAAAAGATGACAGCTTGGCTTGGCTGGCGTGGATAGGC	1148
Db	6102	TTCTATTGTACATATAGGATCTACTGCAAAAGATCAGAAATTAAGTGTCAATATTCGTGGCATAG	6161
OY	1149	CATCTACTTCTATGATCCCATTAATCTACAGAACTTCCACAAAGTTACAGATGTCTTTAAA	1208
Db	6162	CATCTTTTCTTACAGATCCGATTTACTGCTAATAAGCGCTGGGATTCCTACGATACTTTAAA	6221
OY	1209	AGTTAATGAGACTCCGGCAATTCGTCACTACATATACAGGGAACATCATCTTCACAGG	1268
Db	6222	TCTCATTAAGGCTCATCAGATATAGTACAGATTAATATAGTGGGTGGAATGTGTTTTCTTGG	6281
OY	1269	AGAAAAGTTATCAGACAGACAGGCGCAGATTTCTTAAAACTTCTTGGAGAGCTACTACA	1328
Db	6282	TGAAAAGCTCTGTGAATATAGCAAAAAGTGTGAGCAACCTCATCTTCTACGCTGAAGCA	6341
OY	1329	GCCTGTAACTCTTTACAGAGGTACTGTATCTTTAAAAATGAGAGTCACTCTGACAGTCA	1388
Db	6342	GCCTGTAACTCTTAATCTCAGAGAAATTTAGTACTTAAACGTGGTGTCACTCTCGATACGAA	6401
OY	1389	GGCATTCACATCAACAGGCAATTCCTGTCTCCAAATGAGACGTAGACTCTATGAAAC	1448
Db	6402	AGGCTTATCTACAGACCGCGGGTTCCTGTATTATAGATGGGGCAACAGCTTAAAAAGC	6461
OY	1449	---TGCATGTACTAGCAACCAATTAACAATTTGGTCATTAATCAATCACTGTCTATAGAGCGGTGC	1505
Db	6462	AAGTACAGAGGAGGTCACTTTAAACAGTCTTTCCATTTCTGTAGACTCTTTAAGCGGAGG	6521
OY	1506	AAAGAAGGCAAAAATATAGAAACCAAAAGCTACGTCAAAATAATCTGACTTTATCTGAAACAT	1565
Db	6522	TAAAGAAAGTTGTATTTGCTGCTCTGTGAGAGATGAATAAATATAGCCCTTAGTGGTGCAT	6581
OY	1566	CACTTTATTTGAGACCGGAGCGGCACTTTATATGAAATCATATGTTTAAGAAATCTCACTC	1625
Db	6582	TCTTCTTTTGGTATTAACCAAGGGAATGCTTATGAAATACGACTTAGGAAAAACATCAAGA	6641
OY	1626	CTACGACATCTTACAGTCAAAAGCTTCTGGAAGTGTAAACACGACCGAGAGTACCCGA	1685
Db	6642	CTTTTTCATTTGTGTGACGCTCTCTGCTCTGGGTATGCAAAACATTAAGATGTTCACGGCT	6701
OY	1686	TCTCATATATGGGTGAGAAAATTCATTACGGGCTATCAGGAACTTGGGCGCCAAATGTGTTG	1745
Db	6702	TCTCATAGTATGAAACCTCTACGACATATAGGATATCAAGTACTTGGGAAATGACTGGCT	6761
OY	1746	GG-----GGACAGGGGCTCTACAGCTGACATGCAACCTTCAACATGAGATAAACTGG	1793
Db	6762	TGATGATACCGCAAGCACTTCCAAAGACTTAAGACAGAGCATTAAGCTTGTGACCAATACAGG	6821
OY	1794	CTATATTTCTTAATCCGAGGCTATGGGCTTTATAGCCCTAATACCTATAGGAATGCAAT	1853
Db	6822	CTACCTTCCGAATCTCTGAGGCTCAAGGACCTTTAGTCTCTAATTAACCTTTTGGGATCTTT	6881
OY	1854	TATATGATTTAGCTCTCTCCATATATCTTATGAGACGTGCAAAAGAAAGGTTGACAGGGA	1913
Db	6882	TTTCAGACATCCAAAGCAATTTCAAGGTGTCAATAGAGAAAGTGTCTTTGACTCTTTTGTTCAGA	6941
OY	1914	CCGAGCTTTTGGTGTGCTGGATTAATCTAACTCTTCCATAGAGTATGACAAAAACAGC	1973

Db	6942	TCGAGCGTTCGGGCTGCGGAGTCGCCCAATTTCTTAGTAAAGATAAAGAAAGGGGAAAA	7001
Qy	1974	ACGGGGGTTTGGCCATTTGAGTGCGGGTTTGTGCATAGAGGAAACCTACATACTTGTTC	2033
Db	7002	ACGCAAAATACGTCATAAATCTGGTGGATATGCTATCGGAGGTGCAGGCAAACTTGTTC	7051
Qy	2034	AGATATAGTCTTATGTCGTCATTTTGTACGTCCTTGGACAGATATAGACTACTTGT	2093
Db	7062	TGAAAACCTAATATATGCTTTGCTTTTGGCCAACTCTTTGGTACGATAAAGATTTCTTAGT	7121
Qy	2094	AGCTAAGATATCAAGTACAGCTACAGGAGAACTCTATTACAGCAACCAAGAAACCTA	2153
Db	7122	CGCTAAATATATCTATATACCTATGAGAGGACCTTCTATATCCAAACATTTACGAATG	7181
Qy	2154	TATCTCTTCCTTCGCAACATACAGGCGCTGTTCGTGTATTTATGTTTCTACAGAGATTC	2213
Db	7182	TATGCGGTTCATAGTGTGTCTCTTAGATAAA---CTCTCGGCTCTTGGAGTCATTAAC	7288
Qy	2214	TGTTCTCTTTACAGAAACCTTAGCTACACCCATACGATACGATACGATCGAAACCAAGTA	2273
Db	7239	CCTGCTTTTAAAGAGGAGCGCTGCCTTATAGCCACGTCATATATGATCTGAAGACAAGTA	7298
Qy	2274	TACAACATATCTACCTGTTTAAGAGAGCGGGGGAATGATATTCTCGTTAGAATTGCG	2333
Db	7299	TACTGCGTATCTTAGGTGAAGAGTTCTTGGGGAAATAATGCTTTTAACATGATGTGG	7358
Qy	2334	TGGAAGAGCTTCGATTTGCTTAGATGAAAGTGTCTATTTTAGACATACATGCCCTTCAT	2393
Db	7359	AGCTTCTTCATCTTATATCGTAATACGTGACGTATGTTTATGATACCTATAGCTCATACAT	7418
Qy	2394	GAATATGAGTTTCTATATGACATACAGAAAGTTTAAAGAAACGGGAACAGACGCG	2453
Db	7419	CAAAATGATGTACCTATATATAGTCAGACACCTTCTCGGAAAGGTACAGAAAGAG	7478
Qy	2454	TGAATTTGGAAGTACCGCTGTGTGAAATCTGGCTTACCTATACGGGATCCGATTTGTAA	2513
Db	7479	ATCTTTTATATACAGCAACCTCTTCAATTATCTTTGCCCTAATAGGGGTGAAGTTTGAA	7558
Qy	2514	GGAAATCAGACTGCCAAGATCAACAGTACAATCTTGGTTATATCTGTGATGATCTGT	2573
Db	7539	GTTCTCATTTGTAATGACCTTTTCTATATGCTGACTTATATCTATATGTTCCGATCTTAT	7598
Qy	2574	TGCTATGTAACCCGACGTACAGCAACACCTGCGAATTTACGGTGTATCTTGGAAAACCT	2633
Db	7599	CCGCAATATATCCCAAATGACACTACACACTGTAAACACAGGCGCTTTTGGAAACTTA	7658
Qy	2634	CGGTATGCAATTTGGCAGACACACTTATAGCTGTGCGAGGAGCAACATTTTTCCTTTAA	2693
Db	7659	TGCCAATATCTTACACAGACAGGCGTTGCAAGTGGCGTGCAGGCAATCACTACGCTTCTC	7718
Qy	2694	CTCAAAATTTTGAAGCCTTTAGCCAAATTTTCTTTGAAATTCGTGGGTGCATCTGCCAATTA	2753
Db	7719	TCTATATGTTGAATGCTCGGCCAGTTTGTCTTTGAAGTTCGTGATCTCCACGGAATTA	7778
Qy	2754	CAATGTAGACTTAGAGACAAATATCAATTTCTA	2786
Db	7779	TAAATATAGATCTTGGGGTAAAGTTTCAAAATTTTA	7811

RESULT	8
AP002546/C	
LOCUS	
DEFINITION	AP002546 30650 bp DNA BCT 08-DEC-2000 Chlamydomophila pneumoniae J138 genomic DNA, complete sequence, section 2/4.
ACCESSION	AP002546 AB0333782 AB0333783 AB0333784 AB0333785 AB0333800 AB033801 AB033802 AB033803 AB033804 AB033805 AB033806 AB033807 AB033808 AB033809 AB033810 AB033811 AB033812 AB033813 AB033814 AB033815 AB036079 AB036080 AB036081 AB036082 AB038348 AB038349 BA0000001 AP002546.2 GI:10176692
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	Chlamydomophila pneumoniae J138 (strain:J138) DNA.





```

Db 4148 ATTGCAGAAAGATTAAGACTACTGTGAAGCAAGACCGCGAAAGCTGTATGCGGGTTC 4207
QY 2127 TCTGTATTAACAGCA-----CAAGGAACCATATCTCTCTCTCGCAAACTACGCGCC 2180
Db 4208 TGTATATTATGACGCTGACCAAGTTGATGATCTCAGCGGGTATTATTAATGGCGCTAA 4267
QY 2181 TTGTCGCTGCTATGCTATGCTTACAGATTCCTGCTCTTTTCAGGAAACCTTAGCTA 2240
Db 4268 CACGCTGCTTACAGGCTTTTCTAAAGATTCCTATTTCTTGATGACAAATTAACCTA 4327
QY 2241 CACCCATACGATTAACGATCTGAAAACCAAGTATACAAATCTCTACTGTTAAAGAA 2300
Db 4328 TTGGCACACGCGCAACAACAGACACCTCCTATACAGACTATCTGGAAGGATTC 4387
QY 2301 CTGGGGGATATAGTTTTCCTTGAATGATGCGTGAAGAGACTCCGATTTGCTTAGATGA 2360
Db 4388 TTGGGGATATAGTACCCGCGCTTAACTTTGCTTACTAGCGCTACTACCGGATTTAG 4447
QY 2361 AAGTCTCTATTTGACAGTACATGCGCTCATGAAATTTGAGATTTGCTATGCATCA 2420
Db 4448 TTCTCTATCTTTGATAGTATGACACCTTTGCAAAATTAACAGTGTCTATGCGCACCA 4507
QY 2421 GGAAGGTTTTAAAGACAGGAGCAAGAGCTGTAATTTGGAAGTACCGCTTTGTGAA 2480
Db 4508 AGATGACTTTAAAGAACCAACAGAGAGCGCGGCTTTGAAAGCAGCAGATCTTCTCAA 4567
QY 2481 TCTTGCTTACCTACCGGATCCGATTTGATAGGAATCAAGCTGCCAAGATGACAGCTA 2540
Db 4568 CGTTTCTGCTTACCTAGGATTAATAATTTGAGAACTCCCTCATGAGAGAGAGAGTCTTA 4627
QY 2541 CAATCTAATCTTGGTGTACTGTGCTGTAGTACTGTTGCTAGTAAACCGCGCTAGACAGAAC 2600
Db 4628 TGATCTTACACTGATGATATACCTGATGCTGACCGTATATACCACTGATGACAGG 4687
QY 2601 ACTGCGAATTAACGCGTATCTTGGAAAACTTGGTACGAAATTTGGCAGACAGACTTT 2660
Db 4748 CATGATTCGCGCGGATTAACCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4807
QY 2721 TTCTTTGAATGCGTGGGATCTGCAATTAACCTAGTACCTAGTACCTAGTACCTAGTACCT 2780
Db 4808 TGGTTTGAATTAAGAACGCTTTCAGAAATTAATACCTAGTACCTAGTACCTAGTACCT 4867
QY 2781 ATTCTAA 2787
Db 4868 GTTCTAA 4874

```

```

RESULT 7
AE002192 DNA BCT 30-MAY-2000
LOCUS Chlamydomophila pneumoniae AR39, section 25 of 94 of the complete
DEFINITION genome.
ACCESSION AE002192 AE002161
VERSION AE002192.2 GI:8163405
KEYWORDS
SOURCE Chlamydomophila pneumoniae AR39.
ORGANISM Chlamydomophila pneumoniae; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis Mogen and Chlamydia
pneumoniae AR39
JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE 20150255
PUBMED 10684935

```

```

REFERENCE 2 (bases 1 to 12676)
AUTHORS White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE Submitted Submission
JOURNAL Medical Center Dr. Rockville, MD 20850, USA
COMMENT On Jun 1, 2000 this sequence version replaced gi:7189226.
FEATURES
source
location/Qualifiers
1..12676
/organism="Chlamydomophila pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
complement(90..185)
/gene="CP0300"
complement(90..185)
/gene="CP0300"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA03157.1"
/db_xref="GI:7189227"
/transl_table=1
/note="MFTSLGRSAIFVTGSLKHLRLEPLEIFOK"
complement(275..1819)
/gene="CP0301"
complement(275..1819)
/note="CP0301"
/note="similar to GP:4376735; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="polymorphic membrane protein G family"
/protein_id="AA03158.1"
/db_xref="GI:7189228"
/transl_table=1
/note="MFTLRFNLTCSALFLPAAQVYVYHESDGYVAINKSLERK
ITCPEPTSTYTFIDVRIISVAKIDQEDAGVINSGLFPMGNCNTEFNHLCMEG
AITSNRYGDTTFLDVRISVAKIDQEDAGVINSGLFPMGNCNTEFNHLCMEG
SGAITYPYLLGSKASRVSYNLGNRLVRFDRNVSGGAIETHNLTLPKPCPE
NNHAYHDVNSNGALAIAPGSGISIVKSDLLFKPGTAODGNTIHNSIHLSGAF
KNRAVSEGVYVYPPDPSHSHKIDLYVNPAGKRTYPTTSGLCIDDHVCAE
NLTSTIIQDVTLAGGTLISLSDGYTLQHSKQKQASSTLTSPGTTLLCSGDARONH
ILIEDDNEFVPRVIRAEKDALVLEKLVAFAYVSVDFPQKFAETIPLLELGP
SFDLLETFTLETTQVTEENDAVRGFWLSWEYPPSLDKDRITPTKRTVELTNP
EITSIP"
complement(207..4926)
/gene="CP0302"
complement(207..4926)
/note="CP0302"
/note="similar to GP:4376733; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="polymorphic membrane protein G family"
/protein_id="AA03159.1"
/db_xref="GI:7189229"
/transl_table=1
/note="MFTLRFNLTCSALFLPAAQVYVYHESDGYVAINKSLERK
ITCPEPTSTYTFIDVRIISVAKIDQEDAGVINSGLFPMGNCNTEFNHLCMEG
AITSNRYGDTTFLDVRISVAKIDQEDAGVINSGLFPMGNCNTEFNHLCMEG
SGAITYPYLLGSKASRVSYNLGNRLVRFDRNVSGGAIETHNLTLPKPCPE
NNHAYHDVNSNGALAIAPGSGISIVKSDLLFKPGTAODGNTIHNSIHLSGAF
KNRAVSEGVYVYPPDPSHSHKIDLYVNPAGKRTYPTTSGLCIDDHVCAE
NLTSTIIQDVTLAGGTLISLSDGYTLQHSKQKQASSTLTSPGTTLLCSGDARONH
ILIEDDNEFVPRVIRAEKDALVLEKLVAFAYVSVDFPQKFAETIPLLELGP
SFDLLETFTLETTQVTEENDAVRGFWLSWEYPPSLDKDRITPTKRTVELTNP
EITSIP"
complement(207..4926)
/gene="CP0302"
complement(207..4926)
/note="CP0302"
/note="similar to GP:4376733; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="polymorphic membrane protein G family"
/protein_id="AA03159.1"
/db_xref="GI:7189229"
/transl_table=1
/note="MFTLRFNLTCSALFLPAAQVYVYHESDGYVAINKSLERK
ITCPEPTSTYTFIDVRIISVAKIDQEDAGVINSGLFPMGNCNTEFNHLCMEG
AITSNRYGDTTFLDVRISVAKIDQEDAGVINSGLFPMGNCNTEFNHLCMEG
SGAITYPYLLGSKASRVSYNLGNRLVRFDRNVSGGAIETHNLTLPKPCPE
NNHAYHDVNSNGALAIAPGSGISIVKSDLLFKPGTAODGNTIHNSIHLSGAF
KNRAVSEGVYVYPPDPSHSHKIDLYVNPAGKRTYPTTSGLCIDDHVCAE
NLTSTIIQDVTLAGGTLISLSDGYTLQHSKQKQASSTLTSPGTTLLCSGDARONH
ILIEDDNEFVPRVIRAEKDALVLEKLVAFAYVSVDFPQKFAETIPLLELGP
SFDLLETFTLETTQVTEENDAVRGFWLSWEYPPSLDKDRITPTKRTVELTNP
EITSIP"

```



Db 22629 ACCATCATTTATGTGACCCGACGGGACGTTTTATGAAATCATAGTTTAAGAATCT 22688  
Qy 1621 CAGTCCACGACATCTTAGACCTAAAGCTCTGTGAAGTGTAAACAGCAGCCGAGTACT 1680  
Db 22689 CAGTCCACGACATCTTAGACCTAAAGCTCTGTGAAGTGTAAACAGCAGCCGAGTACT 22748  
Qy 1681 CCAGATCCCTAATAGGGGAGAAATTCATATAGCGCTATCAGGGAGCTGGGGCCCAAT 1740  
Db 22749 CCAGATCCCTAATAGGGGAGAAATTCATATAGCGCTATCAGGGAGCTGGGGCCCAAT 22808  
Qy 1741 GTTTGGGGGACAGGGGCTTTCAGCATGCAACCTTCAACTGGACTAAACTGGCTATATT 1800  
Db 22809 GTTTGGGGGACAGGGGCTTTCAGCATGCAACCTTCAACTGGACTAAACTGGCTATATT 22868  
Qy 1801 CCTAATCCCGAGCTATCGGCTCTTATAGTCCCTAATAGCTTATGGAATGCACTTATAGAT 1860  
Db 22869 CCTAATCCCGAGCTATCGGCTCTTATAGTCCCTAATAGCTTATGGAATGCACTTATAGAT 22928  
Qy 1861 ATTAGCTCTCCATATATATCTTATGAGACCTGCAACGAAAGGGTTCAGGGAGCCGTGCT 1920  
Db 22929 ATTAGCTCTCCATATATCTTATGAGACCTGCAACGAAAGGGTTCAGGGAGCCGTGCT 22988  
Qy 1921 TTTTGTGTGCTGATTATCTAATCTTCCATAGGATAGTACAAAAACACAGCGGG 1980  
Db 22989 TTTTGTGTGCTGATTATCTAATCTTCCATAGGATAGTACAAAAACACAGCGGG 23048  
Qy 1981 TTTTGCATTTGATGCGCGGTATGTCATAGAGAAACCTACACTCTTCTTCATGATTAAG 2040  
Db 23049 TTTTGCATTTGATGCGCGGTATGTCATAGAGAAACCTACACTCTTCTTCATGATTAAG 23108  
Qy 2041 ATTTCTATGCTGATTTTGTGACGCTTTTGGAGAGATAGAGACTCTTGTGAGTAAAG 2100  
Db 23109 ATTTCTATGCTGATTTTGTGACGCTTTTGGAGAGATAGAGACTCTTGTGAGTAAAG 23168  
Qy 2101 AATCAAGTACAGTCTACGAGAGAACTCTATTACACAGCAACGAACCTATATCTCT 2160  
Db 23169 AATCAAGTACAGTCTACGAGAGAACTCTATTACACAGCAACGAACCTATATCTCT 23228  
Qy 2161 CTTCCTTGGAAACCTACGCGCTTGTCTGTCTATATGTTCTTACAGAGATTCCTGTTCTC 2220  
Db 23229 CTTCCTTGGAAACCTACGCGCTTGTCTGTCTATATGTTCTTACAGAGATTCCTGTTCTC 23288  
Qy 2221 TTTTCAAGAAACCTTACGCTACACCCATACGGAATGATCTGAACCAACATATACAA 2280  
Db 23289 TTTTCAAGAAACCTTACGCTACACCCATACGGAATGATCTGAACCAACATATACAA 23348  
Qy 2281 TATCTAATCTTAAAGAGAGCTGGGGGAATGATGTTCCCTTGAATTCGTTGAAGA 2340  
Db 23349 TATCTAATCTTAAAGAGAGCTGGGGGAATGATGTTCCCTTGAATTCGTTGAAGA 23408  
Qy 2341 GCTCCGATTTGCTTAGTAAAGAGCTCTATTGAGACATGACCCCTCATGAATTTG 2400  
Db 23409 GCTCCGATTTGCTTAGTAAAGAGCTCTATTGAGACATGACCCCTCATGAATTTG 23468  
Qy 2401 CAGTTTCTATGACATCAGAGAGGTTTTAAAGAACAGGGAACAGAGCTCGTGAATTT 2460  
Db 23469 CAGTTTCTATGACATCAGAGAGGTTTTAAAGAACAGGGAACAGAGCTCGTGAATTT 23528  
Qy 2461 GGAAGTAGCCGCTCTTGAATCTTGGCTTACATATCGGATTCGATTTGATAGGAATCA 2520  
Db 23529 GGAAGTAGCCGCTCTTGAATCTTGGCTTACATATCGGATTCGATTTGATAGGAATCA 23588  
Qy 2521 GACTGCCAATATGACATGATCAATCTTACCTTGTATATCTGATCTTGTTCGTAGT 2580  
Db 23589 GACTGCCAATATGACATGATCAATCTTGTGTATATCTGATCTTGTTCGTAGT 23648  
Qy 2581 AACCCGACATGACAGCAACACTGCGAATTAGCGGTATCTTGGAAACCTTCGTAGC 2640  
Db 23649 AACCCGACATGACAGCAACACTGCGAATTAGCGGTATCTTGGAAACCTTCGTAGC 23708  
Qy 2641 AATTTGGCAAGACAGCTTTAGTCCCTTGTGCAAGGAACATTTTCTTAACTCAAT 2700  
Db 23709 AATTTGGCAAGACAGCTTTAGTCCCTTGTGCAAGGAACATTTTCTTAACTCAAT 23768

Qy 2701 TTGAGCCTTAGCCATTTTCTTGAATTTGGGTGGTATCCGATTTACATGTA 2760  
Db 23769 TTGAGCCTTAGCCATTTTCTTGAATTTGGGTGGTATCCGATTTACATGTA 23828  
Qy 2761 GACTTAGGACCAAAATACCAATTTCTAA 2787  
Db 23829 GACTTAGGACCAAAATACCAATTTCTAA 23855  
RESULT 6  
CPV72499  
LOCUS CPV72499 4926 bp DNA BCT 13-MAY-1999  
DEFINITION Chlamydia abortus putative outer membrane protein gene, partial  
cgs, and putative 98 kda outer membrane protein gene, complete cds.  
ACCESSION U72499  
VERSION U72499.1 GI:1657776  
KEYWORDS  
SOURCE Chlamydia abortus.  
ORGANISM Chlamydia abortus; Chlamydiaceae; Chlamydia.  
REFERENCE 1 (bases 1 to 4926)  
AUTHORS Longbottom, D., Russell, M., Dunbar, S.M., Jones, G.E. and Herring, A.J.  
TITLE 98kda protein genes from ovine abortion strain 526/3 Chlamydia  
psittaci  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4926)  
AUTHORS Longbottom, D.  
TITLE Direct Submission  
JOURNAL Submitted (25-SEP-1996) Moredun Research Institute, 408 Gilmerton  
Road, Edinburgh EH17 7JH, Scotland, UK  
FEATURES  
source  
1..4926  
/organism="Chlamydia abortus"  
/isolate="ovine abortion 526/3"  
/specific\_host="Ovis aries"  
/db\_xref="taxon:83595"  
/clone="5112"  
/dev\_stage="elementary body"  
/note="submitted as Chlamydia psittaci"  
<1..1950  
/codon\_start=1  
/transl\_table=11  
/product="putative outer membrane protein"  
/protein\_id="AAB18187.1"  
/db\_xref="GI:1657777"  
/translation="DPTTTAPRAAAVTAQDEASLAKATSGIPASGSAVYAPAP  
TKPLKINAPDTPDPEIOKVAEAAQOASVYNGKIVFSEKLSSEDAKPLNATVIA  
NDVSLAGTLVLSGGAGLLVDSTFOKESLIVADGCTSLITNTVPASEGLOSTPS  
PKNAIPVIRAVSKVIALSLINLBERADSGAVVPTIESPDGSLITNLAVLDSLE  
NCKVITLAKGGSGSVTLIGDLOPORSQNFVNPILNKNFTLNFIDISAPDAEKIHT  
EGNIIIPGATSSNLGTGCKWEYTEYKDSGKXSFEMKRYSTIETPARRATLVNS  
VWCSAIDMRAPFQWLVSTEGEDFHGGLWISGVSNFHNDSTKVOGFPNHSISGVVG  
VSTQPSINRYMDLAFQMLGSKSDYRIADARSHVYAAIHTKCEKLVNHYTSKRGA  
ILARKPEKSPVLIIDAOISLSHNSMTHTTPRPSRGKMNHCVAEGELSYLPILVD  
HPIARELPFVKLHIVFOEDERKTOGEGENPFSAGFVNVSLDGLAFKETNKL  
TYNIRLAQPDRIADAPKSKVFLPSVHTAMWSTGATNLSDQAMILDSDDHNLIDNLEY  
FCHGAFELGSSSNRYNDIGRKPF"  
2083..2089  
/note="putative"  
2094..4874  
/codon\_start=1  
/transl\_table=11  
/product="putative 98 kda outer membrane protein"  
/protein\_id="AAB18188.1"  
/db\_xref="GI:1657778"  
/translation="MRPSLYKILISLTLPISFHSQTHAEVALFOESILDNGAFS  
POSTSAGTIVNESDISLVQDFOALASSAFVQTAOLTPFGKNNHSLSTNNAAG  
ANPAGINVTADKILILTPDSKLSREKCSLYNTGKAMKSGGALANLANNASILEDQ  
NTSAENGAIISCAFSLTGSKSKEISFTTSTAKGAGATAGTIAHLISDQGTITRESG  
TAVNSGAVYSEASMTIAGNNHVAFSNNVAVSGSDCGGAIHCSKTSGAPTLIRNK  
VILFEENTSSAKGAIYTDKLLITSGPFIINNKYTHATPKGATGAIANGECSLTA  
EHODITFDNNMLATQDNATIKRMAINIEGNGKFNLRASGKTSISFYDPTVEGNAD

```
/gene="CPj0008"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/db_xref="GI:8978381"
/db_xref="GI:8978381"
/transl_table=11
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98219.1"
/db_xref="GI:8978382"
/translation="MMLRYADKFLIREKEEMERHELFHATMYKASGHAYAKAKA
FEKESNENORVKVDEKWLKGLAEPFNOSRRARBEPLRLQTLXPVSVSEVRLER
ORTKRVNENLXADIEKKYHHCVRPDEHYEVEKEKELPYRENGKVLXSAEVSCLIO
RLEDCLEFWSKLLTKAESVFEKMPDATERLGNKVLSDVTRLELLICDAEEMLRLE
ELEMTRVLEPLLEKNTPEKASLOYSCKEMLAKVPOCKESPSTYRSQERLERLN
ODLOATYVNCOEIRLOGFSDESKEVTCSDHIREOKRHEVOGLNFINELLVWGLAE
TOARDLVAIVPMFEFLYOHNIKREKRSOMATERYREIROAFQGMEDLLAED
TLIKEDYWLRLDMLRLDERKNRRLICNKIAAQQRVKGF"
13104..15428
/gene="CPj0010"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/gene="CPj0010"
13104..15428
/gene="CPj0010"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

Query Match      99.5%; Score 2771.8; DB 2; Length 299650;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2784; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1  ATGAAGCTCTTCCTCCCAAGTTGTATTTCTACATTTGGTATTTGCCCTTGTCTATG 60
Db 21070 ATGAAGCTCTTCCTCCCAAGTTGTATTTCTACATTTGGTATTTGCCCTTGTCTATG 21129

QY 61  ATTGCTACCCAGACAGTATTGATTCAGTGCAGATTGATGGAATAAATGGTAAT 120
Db 21130 ATTGCTACCCAGACAGTATTGATTCAGTGCAGATTGATGGAATAAATGGTAAT 21189

QY 121 TTTTCAGTTCGTGAGAGTCAGGAAGATGCTGGAATCTACTACCTATTAAAGGAAATGTC 180
Db 21190 TTTTCAGTTCGTGAGAGTCAGGAAGATGCTGGAATCTACTACCTATTAAAGGAAATGTC 21249

QY 181 ACCTAGAAAATATTCCCTGACAGCAGACAGATCAAAAAGCGTTTAAACAACACT 240
Db 21250 ACCTAGAAAATATTCCCTGACAGCAGACAGATCAAAAAGCGTTTAAACAACACT 21309

QY 241 AAGGCGATTTGACTTTCACAGGTAACGGGAATCTCTATTGTTCCAAAGCGTGATGCA 300
Db 21310 AAGGCGATTTGACTTTCACAGGTAACGGGAATCTCTATTGTTCCAAAGCGTGATGCA 21369

QY 301 GGGACTGTAGCAGGGGCTGCTGTTAAACAGCAGCGTGTAGATTAATCTACACGTTTATA 360
Db 21370 GGGACTGTAGCAGGGGCTGCTGTTAAACAGCAGCGTGTAGATTAATCTACACGTTTATA 21429

QY 361 GGGTTTCTTCGCTATCTTTAFTTGCGTCTCGTGAAGTTCATATACTCCGGCAAAAGA 420
Db 21430 GGGTTTCTTCGCTATCTTTAFTTGCGTCTCGTGAAGTTCATATACTCCGGCAAAAGA 21489

QY 421 GCGGTAGCTGCTACAGGGTACGTTGAAGTGAACAAAATGTCAGTTGCTCTTACAC 480
Db 21490 GCGGTAGCTGCTACAGGGTACGTTGAAGTGAACAAAATGTCAGTTGCTCTTACAC 21548

QY 481 AAAAAGCTTTCAACGATTAATGGCGGTCTATCACCGCAAAAAGCTTTCAATTAACAGGG 540
```

```
|||||
Db 21549 AAAAAGCTTTCAACGATTAATGGCGGTCTATCACCGCAAAAAGCTTTCAATTAACAGGG 21608
QY 541 ACTACAACTGACGCTGCTGTTTCTGAAAATFACCTCCTCAAGAAAAGCGGAGCATTTG 600
Db 21609 ACTACAACTGACGCTGCTGTTTCTGAAAATFACCTCCTCAAGAAAAGCGGAGCATTTG 21668
QY 601 ACTTCGATGCCCTTACCATTTACTGTGAAAACCAAGGGGAGTCTCTTTTCTGCAATFAC 660
Db 21669 ACTTCGATGCCCTTACCATTTACTGTGAAAACCAAGGGGAGTCTCTTTTCTGCAATFAC 21728
QY 661 TCTTGCGATTCGTGAGCTGCAATTTTACAGAAAGCTCGGTGACTATTTCTAATAATGCT 720
Db 21729 TCTTGCGATTCGTGAGCTGCAATTTTACAGAAAGCTCGGTGACTATTTCTAATAATGCT 21788
QY 721 AAGGTTCTCTTATTGACAAATAAGGTCAAGAGAGCGAGCTCCTCAACAAGGGGGATATG 780
Db 21789 AAGGTTCTCTTATTGACAAATAAGGTCAAGAGAGCGAGCTCCTCAACAAGGGGGATATG 21848
QY 781 TCAGAGAGTGTATCTGTGCTTATAAACTAGTACAGATTAAGGTCAACCTCACTGGA 840
Db 21849 TCAGAGAGTGTATCTGTGCTTATAAACTAGTACAGATTAAGGTCAACCTCACTGGA 21908
QY 841 AATCAGATGTACTCTTCAGCAACAATACATGACACAACAGCGGAGAGCTATCTATGTG 900
Db 21909 AATCAGATGTACTCTTCAGCAACAATACATGACACAACAGCGGAGAGCTATCTATGTG 21968
QY 901 AAAAGCTGGAAGCTGGCTCCGAGAGACTTACCCTTTGAGTAAGAAATGTGTCAATGGA 960
Db 21969 AAAAGCTGGAAGCTGGCTCCGAGAGACTTACCCTTTGAGTAAGAAATGTGTCAATGGA 22028
QY 961 GGTACAGCTCTCTAAAGGTGAGAGCATAGTATGSAAGATAGTGGGAATGAGTTATCC 1020
Db 22029 GGTACAGCTCTCTCTAAAGGTGAGAGCATAGTATGSAAGATAGTGGGAATGAGTTATCC 22088
QY 1021 GCCGATAGTGTGACATTTCTTTTATGGAATACAGTACTTCTACTCTCTGCGGAGG 1080
Db 22089 GCCGATAGTGTGACATTTCTTTTATGGAATACAGTACTTCTACTCTCTGCGGAGG 22148
QY 1081 AATAGAAATGATGACACTTAGGAACAGATGCAAAAGATGACAGCTTTGGCTCTGCTGCT 1140
Db 22149 AATAGAAATGATGACACTTAGGAACAGATGCAAAAGATGACAGCTTTGGCTCTGCTGCT 22208
QY 1141 GGTAGAGCCATCTACTTCTATGATCCATTAATCACTACAGATCTTCCACAAGTTACAGAT 1200
Db 22209 GGTAGAGCCATCTACTTCTATGATCCATTAATCACTACAGATCTTCCACAAGTTACAGAT 22268
QY 1201 GTCTTAAAGTTAATGACACTCGCGGAGATTCTGCACCTACATATACAGGAAACATATC 1260
Db 22269 GTCTTAAAGTTAATGACACTCGCGGAGATTCTGCACCTACATATACAGGAAACATATC 22328
QY 1261 TTCACAGGAGAAAAGTTATACAGAGAGAGCGGAGATTCTTAAAAATCTTCTCTGGAAG 1320
Db 22329 TTCACAGGAGAAAAGTTATACAGAGAGAGCGGAGATTCTTAAAAATCTTCTCTGGAAG 22388
QY 1321 CTACTACAGCCTGTAACTCTTTCAGAGGTACTATCTTTAAACATGAGTGACTCTG 1380
Db 22389 CTACTACAGCCTGTAACTCTTTCAGAGGTACTATCTTTAAACATGAGTGACTCTG 22448
QY 1381 CAGACTCAGGCAATTCATCTCAACAGGCAATTCCTGCTGAAAATGAGCTAGAACTACT 1440
Db 22449 CAGACTCAGGCAATTCATCTCAACAGGCAATTCCTGCTGAAAATGAGCTAGAACTACT 22508
QY 1441 CTAGAACCTGTGATACCTAGCAGCATTAACATTTGGTCAATTAACATCAGTCTATAGAC 1500
Db 22509 CTAGAACCTGTGATACCTAGCAGCATTAACATTTGGTCAATTAACATCAGTCTATAGAC 22568
QY 1501 GGTGCAAGAAAGCAAAAATAGAAACCAAAAGCTAGCTCAAAAATCTGACTTATCTGGA 1560
Db 22569 GGTGCAAGAAAGCAAAAATAGAAACCAAAAGCTAGCTCAAAAATCTGACTTATCTGGA 22628
QY 1561 ACCATCATTATTGGACCCGAGCGGCAAGTTTATGAAAATCATAGTTAAGAAATCCT 1620
```

REFERENCE 2 (bases 1 to 299650)  
AUTHORS Shira, M.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/Genbank databases.  
Mutsunori Shira, Yamaguchi University School of Medicine,  
Department of Microbiology, 1-1-1 Minamikogushi, Ube, Yamaguchi  
755-8505, Japan (E-mail: mshira@ipo.cc.yamaguchi-u.ac.jp,  
Tel: 81-836-22-2227, Fax: 81-836-22-2415)  
COMMENT On Aug 31, 2000 this sequence version replaced gi:6172286  
gi:6172286 gi:6172310 gi:6172312 gi:6172314 gi:6172316 gi:6172318  
gi:6172320 gi:6172322 gi:6172324 gi:6635158 gi:6635160 gi:6635162  
gi:6635164 gi:6635166 gi:6635168 gi:6635170 gi:6635172 gi:6547422  
gi:6547426 gi:8547430 gi:8978373.  
AB033780-AB033781, AB033792-AB033799: Submitted (25-Oct-1999)  
AB038345-AB038347: Submitted (14-Feb-2000)  
AB036071-AB036078: Submitted (18-Dec-2000).  
FEATURES  
source  
1. .299650  
Location/Qualifiers  
/organism="Chlamydia pneumoniae J138"  
/strain="J138"  
/db\_xref="taxon:138677"  
/note="synonym: Chlamydia pneumoniae (strain J138)"  
complement(1. .282)  
/gene="Cp10001"  
complement(1. .282)  
/codon\_start=1  
/transl\_table=1  
/product="CT001 hypothetical protein"  
/protein\_id="BAA98211.1"  
/db\_xref="GI:8978374"  
/translation="MLGKIRGLSLVILICALNVLIGITHNKLTIARKCGVSTP  
AQDTYIIGIACVICTLSFPCSKSRSHSDSCSGGCHSHSDKN"  
573. .878  
/gene="gacC"  
573. .878  
/gene="gacB"  
573. .878  
/codon\_start=1  
/transl\_table=1  
/product="Glu-tRNA Gln amidotransferase (C subunit)"  
/protein\_id="BAA98212.1"  
/db\_xref="GI:8978375"  
/translation="MQEFLHREIRILLAKASALQLSEELIQEYQTSLSAVTSMKRA  
LAIEDPADSCESLFMHVNVNVEDREDSTVTSDFRBEFLRNVPESLGLKVPYAVIK"  
895. .2373  
/gene="gata"  
895. .2373  
/gene="gata"  
895. .2373  
/codon\_start=1  
/transl\_table=1  
/product="Glu-tRNA Gln amidotransferase (A subunit)"  
/protein\_id="BAA98213.1"  
/db\_xref="GI:8978376"  
/translation="MYRYSALELAKAVTLCGLTATGVTQHFHRIEAEQGVAFISL  
CKQALAEALIDIKRSGEPGLAGVPGIKDNIHTGLTKTCAVLENTOPPD  
ATVERKKKEKEDIIIGKLNDEFAMGSTLTYSAPFHPHNPMDLSRVPGSSGSAAY  
SAPCPALGSDPGSGSTROPACVGVGFRPSYGVARYGLVAFASLLDITGLANTY  
EDVALMDVFSGDPKDAISREFRDSFNSKLSLEVKVLGVPTFLLEGRLDITRENF  
ESSLAIFEGEGTLVDVLDLISHAVSIYILASAEATNLARDGVARYGYSQAT  
ISOLVDSRGEKGEMRIRILGNVLSAERONVYKKAATVAKIVKAFRAFEKC  
EILAMPVCSAPAEIGIIDLPTVLDTYVAMNLAIVLPALVPSGFSKREGLPLIG  
IIGQGDQDQCVGVSGFQHAQIKQLFSKRYAKSVVLGGQS"  
2370. .3836  
/gene="gabB"  
2370. .3836  
/gene="gabB"  
2370. .3836  
/codon\_start=1  
/transl\_table=1  
/product="(Pet112) Glu tRNA Gln amidotransferase (B  
subunit)"  
/protein\_id="BAA98214.1"  
/db\_xref="GI:8978377"  
/translation="MSAVYADWESVIGLEVHVELNTRASKLFFSSALNRFGEDEPNTNIST

gene  
CDS  
4127. .6895  
/gene="pmp\_1"  
4127. .6895  
/gene="pmp\_1"  
/codon\_start=1  
/transl\_table=1  
/product="polymorphic outer membrane protein G family"  
/protein\_id="BAA98215.1"  
/db\_xref="GI:8978378"  
/translation="MRFSLGFLPVSFTLLSVEDTSLSATITSLPESDFHDSQNA  
ERSYNVQAGVYSLTGDSVSLSNVDSNLAKACEFNTVSGVTFAGNHGLYFNISGCT  
TKRGAVLCDDPQATARFSGESTLSPFIOSGDIREOGCLSKNALMLNLYVREFGN  
OSKTKGATISGANTYTGANDSVSYFNQATPFGAHSSGPIQAVNOAETIRPQNTA  
KNSGGAIVSDGIDIDQNAVLFRENEALFTALGKGAVCCLPFGSSSTPVLVTS  
DNKQVFERNHSLMGGAIVARKLISSGPFLFINNIYANSQNLGALAIDTGEI  
SLSAEKTITFGKRTSLPFLNGIHLQNAKFLQARNQYSEFYDPIFSEADSGTO  
LNIHGDKNKKEGTITLSEKSLANDPDKFTIPQNVMLSGYLVIRKGAETVSK  
FTOSPGSHVLDIGTKLISKEDIALGALIDISLSSSTAAVIRKANTKOISVTD  
SIELSPGNAVEDLRMRNQTPLISLEPGAGSVYTAGDELPSVPHYGKQSNML  
AMTGNKVGEEFMDKINIRPREKBNLPLNLMGNAVYRSLMOYQETHASLQTD  
RGLMIDIGNFEVYASSEDNIRYRHNSSGGLVSNVNIITKHTSMFSLQFEDKAY  
AVSNNEVRYMGLSYLYOTYTLGNIFFYASRNPVANGILSRRLQNLMIPIHLKY  
GHATNDKTDYANFPNPKNSMRNCALIECGSGPLLVFNGALFQAIIPMKQLQVY  
AYOGDFEKTADGRFNGSLTSSVGLRFEPLALSOVVLADFSSYIPDIFRKP  
SCFAAIVISGDSMLVPAHYSRHAIVSGGGRHFNFTFLRCGSLTECPHARNTVI  
NCSKRFK"  
complement(7138. .7293)  
/gene="Cp10006"  
complement(7138. .7293)  
/gene="Cp10006"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAA98216.1"  
/db\_xref="GI:8978379"  
/translation="MQEPLRSALLERLEWLVLLGVPSPEPTRSTPEKDANQPKDSR  
NRTELSL"  
7515. .10169  
/gene="Cp10007"  
7515. .10169  
/gene="Cp10007"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAA98217.2"  
/db\_xref="GI:9956083"  
/translation="MFSGLVIVPLDTSTSLSTSLDDEGNPQMRKRIATLYIAL  
SLIINGAVLLVAVIAGLSVYSIPGAGMACALGCMALAGIDVLLKREVPVLAS  
VTTPPGVSPRGSSITSGADSTIRSLPTVLLDGHGHPQMKRIATLYIVLEPILIA  
SGVLLVAVIPGSLISIPAEAGCALGCMALAGIDVLLKREVPVPAPIPEY  
VIDDIDESIRLQDEAALALRPEESAEGYIKVESHLENNKSLPYGHGLEYE  
KHQIRVRSLSIKMPEFLDIRIFEESEEPFISAKRRLIDLATTVERKILAEOLR  
NNRKAFTSYLQPSIPEKIIDNFEKALMKPILSKSICRPTIIFENHEGVALSLEK  
YKVFLEKVIYRSLQKSYRDIQMSAKMKILHGNPFLENNKTIIMKEHEMESLSS  
NRYVFLALSPENVDPSPDKKNDLSGIPCRDLSLSEISRDQOKRAHLKHOSIYTO  
ARDRLDQSKENOKELERAEQETISSWEVKKFEIRVORIOAIQKLPNLERE  
EYTGQETVYPTVOGTASDVTILGITEVSSHEDNONQSVKVLSSHEVSWEK  
OEYGPKKKEPQDQMSLERFTEHIEBELVQADYSKLSYFKVNNKKEVOYAKFEL  
KYLESDLEGILAQTESASLITQBELPILATRLAEKAVFGKSLCAKAKAPYE  
DPRRODPTOLRALTLRLQPAKASLEIEIRFSLNLEADIEERLIESKOTERACL  
GVLEIAVESTYDLRSLTNTWEGTPSEKXYFYMWLYNYNEKRRATRLVEMTQRR  
DFKMALEAMQFNALQBELSTQABSE"  
10645. .11358  
/gene="Cp10008"  
10645. .11358  
/gene="Cp10008"



QY 961 GGTACAGCTCTAAAGGTGAGCATAGCTATCGAAGATAGTGGGAATTGATTATCC 1020  
 DB 8151 GGTACAGCTCTAAAGGTGAGCATAGCTATCGAAGATAGTGGGAATTGATTATCC 8092  
 QY 1021 GCCGATAGTGTGACATTTCTTTTAGGGAATPACATGCTCTTACTACTCTGGAGC 1080  
 DB 8091 GCCGATAGTGTGACATTTCTTTTAGGGAATPACATGCTCTTACTACTCTGGAGC 8032  
 QY 1081 AATGAGATAGTATGACACTTAGAAGAGTGAAGATGACAGCTTGGCTCTGCTGCT 1140  
 DB 8031 AATGAGATAGTATGACACTTAGAAGAGTGAAGATGACAGCTTGGCTCTGCTGCT 7972  
 QY 1141 GGTAGAGCATCTACTTCTATGATCCCATACAGATCTTCCACACACAGTACAGAT 1200  
 DB 7971 GGTAGAGCATCTACTTCTATGATCCCATACAGATCTTCCACACACAGTACAGAT 7912  
 QY 1201 GCTTAAAGTTAATGAGACTCCGGCAGATTTGCACTTCAATATPACAGGGAACATATC 1260  
 DB 7911 GCTTAAAGTTAATGAGACTCCGGCAGATTTGCACTTCAATATPACAGGGAACATATC 7852  
 QY 1261 TTCCACAGAGAAAGTTATCAGAGACAGAGCCGAGATTTCAAAAATCTTACTTCCGAG 1320  
 DB 7851 TTCCACAGAGAAAGTTATCAGAGACAGAGCCGAGATTTCAAAAATCTTACTTCCGAG 7792  
 QY 1321 CTACTACAGCCCTGTAACTCTTTCAGAGGACTCTATCTTAAACATGAGATGACTCTG 1380  
 DB 7791 CTACTACAGCCCTGTAACTCTTTCAGAGGACTCTATCTTAAACATGAGATGACTCTG 7732  
 QY 1381 CAGACTCAGAGCACTTCACTCAACAGGCAATTTCTGCTTCGAAATGAGACTAGAACTACT 1440  
 DB 7731 CAGACTCAGAGCACTTCACTCAACAGGCAATTTCTGCTTCGAAATGAGACTAGAACTACT 7672  
 QY 1441 CTAGAAGCTCTGATAGTACTAGCAACATTAACATTTGGTCAATPACATGCTCTATAGAC 1500  
 DB 7671 CTAGAAGCTCTGATAGTACTAGCAACATTAACATTTGGTCAATPACATGCTCTATAGAC 7612  
 QY 1501 GGTGCAAGAGGCAAAATPAGAAACCAAGCTACGTCAAAATCTGACTTATCTGGA 1560  
 DB 7611 GGTGCAAGAGGCAAAATPAGAAACCAAGCTACGTCAAAATCTGACTTATCTGGA 7552  
 QY 1561 ACCATCAGCTTATGAGACCCGAGGCGACTTTATGAAATCATAGTTAAGAAATCCT 1620  
 DB 7551 ACCATCAGCTTATGAGACCCGAGGCGACTTTATGAAATCATAGTTAAGAAATCCT 7492  
 QY 1621 CAGTCTCAGACATCTTAGAGCTCAAAAGCTTGAAGCTGAACAGCAGCCAGTACT 1680  
 DB 7491 CAGTCTCAGACATCTTAGAGCTCAAAAGCTTGAAGCTGAACAGCAGCCAGTACT 7432  
 QY 1681 CCAGATCTATTAATGAGGAGAAATTCATTAAGGCTATCAGGGAATTTGGGCCCAAT 1740  
 DB 7431 CCAGATCTATTAATGAGGAGAAATTCATTAAGGCTATCAGGGAATTTGGGCCCAAT 7372  
 QY 1741 GTTTGGGGGACAGGGGCTCTACGACTGCAACCTTCAACTGACATAAACTGGCTATATT 1800  
 DB 7371 GTTTGGGGGACAGGGGCTCTACGACTGCAACCTTCAACTGACATAAACTGGCTATATT 7312  
 QY 1801 CCTAATCCGAGACGATCGGCTCTTATAGTCCCTAATAGCTTATGGAATGCAATTTATAGAT 1860  
 DB 7311 CCTAATCCGAGACGATCGGCTCTTATAGTCCCTAATAGCTTATGGAATGCAATTTATAGAT 7252  
 QY 1861 ATTACTCTCTCATTTACTTATGAGAGCTGCAAAAGGAGGTTGACAGGAGAACCTGCT 1920  
 DB 7251 ATTACTCTCTCATTTACTTATGAGAGCTGCAAAAGGAGGTTGACAGGAGAACCTGCT 7192  
 QY 1921 TTTTGGTGTGATATATCTTCACTTCCATTAAGATAGTACAAAACAGAGCGCGG 1980  
 DB 7191 TTTTGGTGTGATATATCTTCACTTCCATTAAGATAGTACAAAACAGAGCGCGG 7132  
 QY 1981 TTTTGGCATTTAGTGGCGGTTATGTCATAGAGAGAAACCTCATACTTGTTCAGATAG 2040  
 DB 7131 TTTTGGCATTTAGTGGCGGTTATGTCATAGAGAGAAACCTCATACTTGTTCAGATAG 7072  
 QY 2041 ATTCTTAGTGTGATTTTGTCAAGCTCTTGGAGAAGATAGAGACTTCTTGTAGCTAAG 2100

DB 7071 ATTCTTAGTGTGATTTTGTGAGCTCTTGGAGAAGATAGAGACTTCTTGTAGCTAAG 7012  
 QY 2101 AATCAAGTACAGCTCTAGGAGAACTCTATTTACAGACACAAAGAAACCTATATCT 2160  
 DB 7011 AATCAAGTACAGCTCTAGGAGAACTCTATTTACAGACACAAAGAAACCTATATCT 6952  
 QY 2161 CTTCCTTGAACCTACAGGCTTGTGCTTATAGTCTTACAGAGATTCCTGTCTC 2220  
 DB 6951 CTTCCTTGAACCTACAGGCTTGTGCTTATAGTCTTACAGAGATTCCTGTCTC 6892  
 QY 2221 TTTTGAAGAACTTACCTACACCCATACGATTAACGATCTGAAAAACAGTATACACA 2280  
 DB 6891 TTTTGAAGAACTTACCTACACCCATACGATTAACGATCTGAAAAACAGTATACACA 6832  
 QY 2281 TATCTACTGTTAAAGAAAGCTGGGGAATGATAGTTTGGCTTGAATTCGGTGAAGA 2340  
 DB 6831 TATCTACTGTTAAAGAAAGCTGGGGAATGATAGTTTGGCTTGAATTCGGTGAAGA 6772  
 QY 2341 GCTCCGATTTGCTTATGATGAAGAGTCTATTTAGAGACATGATCCCTTCATGAATTTG 2400  
 DB 6771 GCTCCGATTTGCTTATGATGAAGAGTCTATTTAGAGACATGATCCCTTCATGAATTTG 6712  
 QY 2401 CAGTTTGTCTATGACACATCAGGAAGTTTAAAGAACAGGAAACAGACCTGCTGAATTT 2460  
 DB 6711 CAGTTTGTCTATGACACATCAGGAAGTTTAAAGAACAGGAAACAGACCTGCTGAATTT 6652  
 QY 2461 GGAAGTACGCGCTTGTGAATCTGCTTACCTATCGGATCCGATTTGATAGGAATTA 2520  
 DB 6651 GGAAGTACGCGCTTGTGAATCTGCTTACCTATCGGATCCGATTTGATAGGAATTA 6592  
 QY 2521 GACTCCGAAGATGACAACTGATCAATCTGCTTATGATGATGATGATGATGATGATGAT 2580  
 DB 6591 GACTCCGAAGATGACAACTGATCAATCTGCTTATGATGATGATGATGATGATGATGAT 6532  
 QY 2581 AACCCGACTGATGACAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640  
 DB 6531 AACCCGACTGATGACAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 6472  
 QY 2641 AATTTGGCAAGCAAGCTTATGCTTGTGCGAGGAAACATTTTGGCTTATGATCAAT 2700  
 DB 6471 AATTTGGCAAGCAAGCTTATGCTTGTGCGAGGAAACATTTTGGCTTATGATCAAT 6412  
 QY 2701 TTTGAAGCTTATGACCAATTTCTTTGAATTTGCGTGGTCTATCGCAATTAATTAAT 2760  
 DB 6411 TTTGAAGCTTATGACCAATTTCTTTGAATTTGCGTGGTCTATCGCAATTAATTAAT 6352  
 QY 2761 GACTTAGAGCAAAATPACCAATTTCTAA 2787  
 DB 6351 GACTTAGAGCAAAATPACCAATTTCTAA 6325  
 QY 2787 GACTTAGAGCAAAATPACCAATTTCTAA 2787  
 DB 6325 GACTTAGAGCAAAATPACCAATTTCTAA 6325

**RESULT 5**  
 AP002545 299650 bp DNA BCT 08-DEC-2000  
 Chlamydia pneumoniae J138 genomic DNA, complete sequence,  
 section 1/4.  
 AP002545 AB033780 AB033781 AB033792 AB033793 AB033794 AB033795  
 AB033796 AB033797 AB033798 AB033799 AB036071 AB036072 AB036073  
 AB036074 AB036075 AB036076 AB036077 AB036078 AB038345 AB038346  
 AB038347 BA000008  
 AP002545.2 GI:9956082  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Chlamydia pneumoniae J138 (strain:J138) DNA.  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE  
 AUTHORS  
 Shiral,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,  
 Shiba,T., Ishii,K., Hattori,M., Kuwata,S. and Nakazawa,T.  
 Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CWI029 from USA  
 Nucleic Acids Res. 28 (12), 2311-2314 (2000)  
 JOURNAL MEDLINE  
 20330349







[illegible]

	Query Match	99.5%;	Score 2773.4;	DB 2;	Length 26920;
	Best Local Similarity	99.9%;	Pred. No. 0;		
	Matches 2785;	Conservative	0;	Mismatches 1;	Indels 1; Gaps
QY	1	ATGAACTCTCTCTTCCCACAGTTTGTATTTCTACATTTGGATTTTGCCTTGTCTATG	60		
DB	17762	ATGAACTCTCTCTTCCCACAGTTTGTATTTCTACATTTGGATTTTGCCTTGTCTATG	17821		
QY	61	ATTGCTACCGAGACAGTATTGGATTTCAATGCGAGTTTGCATGGGAATAAAAATGTAAT	120		
DB	17822	ATTGCTACCGAGACAGTATTGGATTTCAATGCGAGTTTGCATGGGAATAAAAATGTAAT	17881		
QY	121	TTTTAGATTCTGAGAGTCAGGAAGATGCTGGAACCTACCTATTATTAAGGAAATGTC	180		
DB	17882	TTTTAGATTCTGAGAGTCAGGAAGATGCTGGAACCTACCTATTATTAAGGAAATGTC	17941		
QY	181	ACTCTAGAAAATATTTCTCTGGAAACAGGACACAGAAATACAAAAGCTGTTTAACAACT	240		
DB	17942	ACTCTAGAAAATATTTCTCTGGAAACAGGACACAGAAATACAAAAGCTGTTTAACAACT	18001		
QY	241	AAGGGGATTTTGACTTTCACAGATTAACGGGAATCTCTATTGTCTCAAAAGGTGATGCA	300		
DB	18002	AAGGGGATTTTGACTTTCACAGATTAACGGGAATCTCTATTGTCTCAAAAGGTGATGCA	18061		
QY	301	GGGACTGTAGCAGGGGGCTCTGTTAAACAGCAGCGTGTAGATTAATCTACACAGTTTATA	360		
DB	18062	GGGACTGTAGCAGGGGGCTCTGTTAAACAGCAGCGTGTAGATTAATCTACACAGTTTATA	18121		
QY	361	GGGTTTCTTCGGCTATCTTTTATTTGGGTCTCTGGAAAGTTTCATATCTACCGGCAAAAGA	420		
DB	18122	GGGTTTCTTCGGCTATCTTTTATTTGGGTCTCTGGAAAGTTTCATATCTACCGGCAAAAGA	18181		
QY	421	GGCGTAGTGGCTCTACGGGGTACGTGAAGTTTGAACAAAATGTCAAGTTTGCCTTCACG	480		
DB	18182	GGCGTAGTGGCTCTACGGGGTACGTGAAGTTTGAACAAAATGTCAAGTTTGCCTTCACG	18241		
QY	481	AAAACTTTTCAACGGATTAATGGCGGTGCTATCACCGCAAAAACCTTTTCATTAACAGGG	540		
DB	18241	AAAACTTTTCAACGGATTAATGGCGGTGCTATCACCGCAAAAACCTTTTCATTAACAGGG	18301		
QY	541	ACTACATGTCAAGCTCTGTTTTCTGAAAATACCTCTCAAGAAAAGCGGAGCCATTCAG	600		
DB	18301	ACTACATGTCAAGCTCTGTTTTCTGAAAATACCTCTCAAGAAAAGCGGAGCCATTCAG	18361		
QY	601	ACTTCGAGTCCCTTACCACTTACTGTAAGCAAGGGAAGTCTTTTTCTGACAAATCT	660		
DB	18361	ACTTCGAGTCCCTTACCACTTACTGTAAGCAAGGGAAGTCTTTTTCTGACAAATCT	18421		
QY	661	TCTTGGGATTCGAGAGCTGCAATTTTTCAGAGAAGCTCGGTGACTATTCTTAATATGCT	720		
DB	18421	TCTTGGGATTCGAGAGCTGCAATTTTTCAGAGAAGCTCGGTGACTATTCTTAATATGCT	18481		
QY	721	AAAGTTTCTTTATTTGACAAATTAAGTCTACAGAGCGAGTCTCTCAACAACGGGGATATG	780		
DB	18481	AAAGTTTCTTTATTTGACAAATTAAGTCTACAGAGCGAGTCTCTCAACAACGGGGATATG	18541		
QY	781	TGAGGAGGTGCTATCTGCTCTTATTAATAACTATGATACAGATCTTAAGGTCAACCTCACTGGA	840		
DB	18541	TGAGGAGGTGCTATCTGCTCTTATTAATAACTATGATACAGATCTTAAGGTCAACCTCACTGGA	18601		
QY	841	AATCAGATTTTCTCTTTCAGCAACAATACATCATGAGCAACAGCGGAGAGACTATCTATGTG	900		
DB	18601	AATCAGATTTTCTCTTTCAGCAACAATACATCATGAGCAACAGCGGAGAGACTATCTATGTG	18661		
QY	901	AAAAAGCTCGAACTGGCTTCGGAGAGCTTACCTTAATTCAGTGAATATGTCTCAATGGA	960		
DB	18661	AAAAAGCTCGAACTGGCTTCGGAGAGCTTACCTTAATTCAGTGAATATGTCTCAATGGA	18721		
QY	961	GGTACAGCTCCCTAAAGGTGAGACCAATAGCTATGCAAGAAATAGTGGGAAATGTAGTTATCC	1020		
DB	18721	GGTACAGCTCCCTAAAGGTGAGACCAATAGCTATGCAAGAAATAGTGGGAAATGTAGTTATCC	18781		

QY	2561	AA	CCCCGACCTTACGACACACCTGGCAATTCAGCGGTATTTCTTGGAAAAACCTTCGGTACG	2640
Db	2751	AA	CCCCGACCTTACGACACACCTGGCAATTCAGCGGTATTTCTTGGAAAAACCTTCGGTACG	2810
QY	2641	AA	TTTGGCAACAACAGCTTTCAGTTCCTTCGTCAGGAGCAACATTTTTCGTTTAACCAAT	2700
Db	2811	AA	TTTGGCAACAACAGCTTTCAGTTCCTTCGTCAGGAGCAACATTTTTCGTTTAACCAAT	2870
QY	2701	TT	TTCGAGCCTTTAGCCAAATTTCTTTTGAATTCGCTGGGTCAATCTCGCAATTTACATGTA	2760
Db	2871	TT	TTCGAGCCTTTAGCCAAATTTCTTTTGAATTCGCTGGGTCAATCTCGCAATTTACATGTA	2930
QY	2761	GA	CTTACGACGCAAAATACCAATTCCTAA	2787
Db	2931	GA	CTTACGACGCAAAATACCAATTCCTAA	2957
RESULT 3				
LOCUS	CPN133035			
DEFINITION	CPN133035	26920 bp	DNA	BCT 27-FEB-2001
ACCESSION	CPN133035			
VERSION	AIJ133035			
KEYWORDS	AIJ133035	GI:4455890		
SOURCE	H81 gene; H82 gene; h83 gene; h84 gene; h85 gene; h86 gene; h87 gene; omp15 gene; omp6 gene; omp7 gene; omp8 gene; omp9 gene; outer membrane protein 6; outer membrane protein 7; pseudogene.			
ORGANISM	Chlamydia pneumoniae			
REFERENCE	1 (bases 1 to 26920)			
AUTHORS	Daugaard, L., Hjerno, K., Knudsen, K., Madsen, A.S., Christiansen, G.			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 26920)			
AUTHORS	Boesen, T.			
TITLE	Direct Submission			
JOURNAL	Submitted (21-JAN-1999) Boesen T., Department of Medical Microbiology and Immunology, University of Aarhus, The Bartholin Building, DK-8000 Aarhus, DENMARK			
FEATURES				
SOURCE	location/Qualifiers			
	1. 26920			
	/organism="Chlamydia pneumoniae"			
	/strain="VR1310"			
	/db_xref="taxon:83558"			
	498..3266			
	/gene="omp6"			
	498..575			
	/gene="omp6"			
	498..3266			
	/gene="omp6"			
	/note="putative"			
	/codon_start=1			
	/transl_table=11			
	/product="outer membrane protein 6"			
	/protein_id="CAB37075.1"			
	/db_xref="GI:4455891"			
	/db_xref="SPTREMBL:Q924H9"			
	/translation="MRSLICGLPIVLESLTSLVSDTSLASTTSLPEDEPHSGDONNAEERAYVOAGDVSITGLGVSTISVNDNSALNACFEVTSVTFACNHHGILFNNISGSTRKAGVLCODPQATARFSGSPSLFQSGPDICEQCLSKNALNNTVYAFEDONOSTKRGALICANVTYIGRTDYSFVYONAAITFGALISGFLQIAVQALTRAEONAKNSGGLYSDGIDIDQNAVLFRENEALTTAAGKGAACCEPSSVPPVIVTFSDNQALVERNHSIMGGAIYARKLSISSGPTLEINNIYANSNOLGALAIIDGGIISLAENKCTITFOGNRTSLPFLNGLHLLQNKFLQARNCSEIEYDPTSEADGSTQGLNLNAGDCKREKYEYGTILFSGEKSILANDPDPKSTIPQNVNLSAGELYIRKGAETYSKFTFQSPGSHVLIDGTLKAKEDATLGLAIDIDLSLSSSTAAYITKANTKOISYNDSTLISPTGNAYEDIPRHRNSCTPLSLSEPGAGSYVTAGDPLPSPHPIHGOSNMKLTAMTGCKAKVEFFMDKINIRPREKESNLVNLIMGAVANVRSLMVOETHRASLQTDKGLMIDIGNEFHVASSEDNIYRKHNNSGGVVLVNNNITPRHYSMAFSQLFSDKDYAVSNNEIRMLVIGSLYQYTTSLGNIIFPYASRNPNVANGIISRELQNDLMIHFELCAY			
gene				
sig_peptide				
CDS				

gene  
3976. .6870  
/gene="hb1"  
3976. .6870  
/gene="hb1"  
/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB1 protein"  
/protein\_id="CAB37076.1"  
/db\_xref="GI:4455892"  
/db\_xref="SPTREMBL:Q923C8"  
/translation="MRKLRLAIVLAIIVLALSTILNIGVVLTVLAIPLGLSVISYPAGKGC  
ACALICWALGIDVLAKKREVPVLAIVSVTTPTGSPRSISISGADSTIRSLPYTL  
LDEGHPOSKRKLRLAIVLIVFISIIILASGVLTVAIPGLSVISYPAGKGCALIC  
WALGIDVLAKKREVPVLAIVSVTTPTGSPRSISISGADSTIRSLPYTLDEGAP  
OSMRKLRLAIVLIVFISIIILASGVLTVAIPGLSVISYPAGKGCALICWALG  
IDVLAKKREVPVLAIVSVTTPTGSPRSISISGADSTIRSLPYTLDEGAP  
VSHLEKRSKSPYDGGHLEKTHQIRVRSKLAMPPEFLDIRIFDEEFPFLSKR  
KRLIDALTVEERKILTEQLERNNLKRAFSYLYODSIFPKLIIDPEKLAEMFLTSLK  
ICRPIITFEENHGYAKSLILKHNAYLKEAFVYRDSLOKSYRDIGKSSAKMILHNPFF  
SLDEDMKQITIMEKHAEMLESLSYRKVFALSDENVVDTSPDKKMDISGJPCRBALSE  
ISRDQOMKQITIMEKHAEMLESLSYRKVFALSDENVVDTSPDKKMDISGJPCRBALSE  
EVSORIRAIQKLYNIIILREBEETGOETVPTVQGTASSDLDTIIGRIEVSRENN  
ONQESCVIARLSHEVEMSEVKQETGPKKRFQOQKSGLEFFTHIELEIVLQKDS  
KILSTFKVYNNKKEVOYIAKFRKLVESDLDEGILQOTSABESLLOEBELPIATGALAE  
KAVFEGSLSCALASAKAPFEEDPPEQSDDTQLATLRLIQEAKASLEEKRSFMLE  
NDIAERRLKSKQTFERRAGVGLRELAIVESTYDLKSLTNWTWGTEPSEKRVYFSL  
NYYNEEKRAKTRIVEMTORVYRDEKMALEAMQFNEALLOEISTQAPSE"  
7150. .8058  
/gene="hb2"  
7150. .8058  
/gene="hb2"  
/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB2 protein"  
/protein\_id="CAB37077.1"  
/db\_xref="GI:4455893"  
/db\_xref="SPTREMBL:Q925Q3"  
/translation="MQVFLSPQLPEPPRRSLGVSCSKRLSISITLVLVGLVLTGLIP  
GLTGISAGAGIGFSAAGVLVLSGLLEFLVLRREVPVYRSEIIRGVSVSSEPALAE  
KQKPEPKKILDIRLPEKLEDDIOTYIOGFVLEKRLKQKEDGDLTEAKERYLSD  
VYERKMESEPLDIORVLENEAAYEOHODCPLENTAIVFSSQEDRDVQYGVGYLPS  
GDARDRKLRSKVEYMDRFRMYTKMSWASVMDHSIGVANELKRAVGLAESVYVIL  
LFKSTRDYAECEKAKIQRDGRFKWL"  
8185. .9492  
/gene="hb3"  
8185. .9492  
/gene="hb3"  
/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB3 protein"  
/protein\_id="CAB37078.1"  
/db\_xref="GI:4455894"  
/db\_xref="SPTREMBL:Q923C0"  
/translation="WMIDRYADKFLIREKEKEMERHFLFHTATWRKASGAAVAKAKA  
PEKESNNQKRVKVEKWLISKGLAEFRNOSRRARERLRLOTLIYEVYSEBVLDER  
OCTKRVLENLTADILEKTHHCVRQEHYKVEENKAEETRENEKVLSEAVESQLQ  
RLIEDLEWMSKILTKAESVFEKMPDATEKLGKLVSDVYNNRLIICLEDAEMIFRIE  
ELIEMIRAVVELPLLEPMKNTPEKASLOYKSEKMLAKVPOCKSPITYRSQOEIRLERIN  
ODLOIYVANNCOBERLQFSDLESKVRTCDHLREQMKHFEVQKGLFNIIEELLVWGAELF  
TQARDAIVATYVMEFYLOYINIKREKRVSGMMAKTETRYEIRIOAFQVAKMEDLLAD  
TIKREDYVWLIRDMDLNDRKRNQRRILICKIIAAQQRYKGF"  
9805. .10698  
/gene="hb4"  
9805. .10698  
/gene="hb4"  
/note="hypothetical"

```
|||||
Db 532 GGGTTTCGTGCTATCTTTATTGCGCTCTCGGAAGTGTGATACATACCGGAAAGGA 591
QY 421 GCCGTTAGCTCTCTACGCGTAGCTTGAGTTTGACAAAATGTCAGTTTGGCTTTCACG 480
Db 592 GCCCTTACTCTCTACGCGTAGCTTG - AGTTGCAAAAATGTCAGTTTCTCTTACG 650
QY 481 AAAAAGCTTTCAAGGATTAATGGCGGTCTATCACCGCAAAAGCTTCAATTACAGG 540
Db 651 AAAAAGCTTTCAAGGATTAATGGCGGTCTATCACCGCAAAAGCTTCTTCAATTACAGG 710
QY 541 ACTCAATATGCTGCTGTCTTTTGTGAAAATACCTCTCCAAAAGAAAAGCGGAGCCATTAG 600
Db 711 ACTCAATATGCTGCTGTCTTTTGTGAAAATACCTCTCCAAAAGAAAAGCGGAGCCATTAG 770
QY 601 ACTTCCGATGCGCTTACCATTAATGGAACCAAGGGAAGTCTTTTCTGACATATCT 660
Db 771 ACTTCCGATGCGCTTACCATTAATGGAACCAAGGGAAGTCTTTTCTGACATATCT 830
QY 661 TCTTGGATTCGAGCTGCAATTTTTCAGAGCGTGGGACTATTTCAATATGCT 720
Db 831 TCTTGGATTCGAGCTGCAATTTTTCAGAGCGTGGGACTATTTCAATATGCT 890
QY 721 AAAAGTTCTTTATTTGACAAATAGGTCACAGAGCGAGCTCTCAACAGCGGGGATATG 780
Db 891 AAAAGTTCTTTATTTGACAAATAGGTCACAGAGCGAGCTCTCTCAACAGCGGGGATATG 950
QY 781 TCAGAGAGTGTATCTGTCTTATTTAAATCTAGTACATATTAAGTACCTCTACTGGA 840
Db 951 TCAGAGAGTGTATCTGTCTTATTTAAATCTAGTACATATTAAGTACCTCTACTGGA 1010
QY 841 AATAGATATGTTACTCTTCAGCAACAATATCATGCAACAAGCGGGAGGAGTATCATATG 900
Db 1011 AATAGATATGTTACTCTTCAGCAACAATATCATGCAACAAGCGGGAGGAGTATCATATG 1070
QY 901 AAAAAGCTCGAAGTGGCTTCGAGAGACTTACCTATTCAGTAGAAAATATGTCATATGGA 960
Db 1071 AAAAAGCTCGAAGTGGCTTCGAGAGACTTACCTATTCAGTAGAAAATATGTCATATGGA 1130
QY 961 GGTACAGCTCTTAAGGAGGAGCCATAGCTATCGAAGATATGAGGGAATTAAGTTATCC 1020
Db 1131 GGTACAGCTCTTAAGGAGGAGCCATAGCTATCGAAGATATGAGGGAATTAAGTTATCC 1190
QY 1021 GCCGATATGAGTGTGACATTTGTTTAAAGGAATATGACATCTTCTACTACTCTGGGAGC 1080
Db 1191 GCCGATATGAGTGTGACATTTGTTTAAAGGAATATGACATCTTCTACTACTCTGGGAGC 1250
QY 1081 AATAGAGTATGATCGACTTATAGAACAGAGTGCAAAAGATGACAGCTTTCGTTCTGCTCT 1140
Db 1251 AATAGAGTATGATCGACTTATAGAACAGAGTGCAAAAGATGACAGCTTTCGTTCTGCTCT 1310
QY 1141 GGTAGAGCCATCTACTCTATGATCCCATTAATCTACAGAGATTTCCACACAGTTACAGAT 1200
Db 1311 GGTAGAGCCATCTACTCTATGATCCCATTAATCTACAGAGATTTCCACACAGTTACAGAT 1370
QY 1201 GTCTTAAAGTAAATGAGACTCGGCGAGATTTCTGACATACAAATATACAGGAAATATTC 1260
Db 1371 GTCTTAAAGTAAATGAGACTCGGCGAGATTTCTGACATACAAATATACAGGAAATATTC 1430
QY 1261 TTCACAGAGAAAAGTATACAGAGAGAGCCAGATTTCTTAAAAAATCTTACTTCCGAAAG 1320
Db 1431 TTCACAGAGAAAAGTATACAGAGAGAGCCAGATTTCTTAAAAAATCTTACTTCCGAAAG 1490
QY 1321 CTACTACAGCTGTAACTCTTTCAGAGAGTACTATCTTTTAAACATGAGTACTCTG 1380
Db 1491 CTACTACAGCTGTAACTCTTTCAGAGAGTACTATCTTTTAAACATGAGTACTCTG 1550
QY 1381 CAGAGTCAGGATTCATCAACAGAGAGATTCGTCTCGAAATGAGAGCTGGAAGTACT 1440
Db 1551 CAGAGTCAGGATTCATCAACAGAGAGATTCGTCTCGAAATGAGAGCTGGAAGTACT 1610
QY 1441 CTAGAACCTGCTGATACAGACCATTAACAAATTTGGTCATTATACATCAGTTCTATAGAC 1500
|||||
|||||
Db 1611 CTAGAACCTGCTGATACAGACCATTAACAAATTTGGTCATTATACATCAGTTCTATAGAC 1670
QY 1501 GGTCAAAAGAGGCAAAAATTAAGAACCAAGGCTACGTCAAAAAATCTGACTTTATCTGGA 1560
Db 1671 GGTCAAAAGAGGCAAAAATTAAGAACCAAGGCTACGTCAAAAAATCTGACTTTATCTGGA 1730
QY 1561 AACATCATTTTATTTGAGACCCGACGCGGACGTTTATGAAAATCATAGTTTAAAGAAATCT 1620
Db 1731 AACATCATTTTATTTGAGACCCGACGCGGACGTTTATGAAAATCATAGTTTAAAGAAATCT 1790
QY 1621 CAGTCTACAGACATCTTAGAGCTCAAGACTTCTGGAAGTATACAGACCGCAGTACT 1680
Db 1791 CAGTCTACAGACATCTTAGAGCTCAAGACTTCTGGAAGTATACAGACCGCAGTACT 1850
QY 1681 CCAAGTCTTATTAATGAGGAGAAAATTCATAGGCTATCAGAGGAACTTGGGGCCCATTT 1740
Db 1851 CCAAGTCTTATTAATGAGGAGAAAATTCATAGGCTATCAGAGGAACTTGGGGCCCATTT 1910
QY 1741 GTTTGGGGAGAGGGGCTTCTACGACTCAACCTTCAACTGAGTAAAACTGGCTATATTT 1800
Db 1911 GTTTGGGGAGAGGGGCTTCTACGACTCAACCTTCAACTGAGTAAAACTGGCTATATTT 1970
QY 1801 CCTAATCCGAGCGTATGCGCTCTTTAGTCCCTAATATAGCTTATGGAATGATTTATAGAT 1860
Db 1971 CCTAATCCGAGCGTATGCGCTCTTTAGTCCCTAATATAGCTTATGGAATGATTTATAGAT 2030
QY 1861 ATTAGCTCTCTCCATTAATCTTATGAGAGCTGCAAAACGAAGGTTGCCAGGAGCCGTCT 1920
Db 2031 ATTAGCTCTCTCCATTAATCTTATGAGAGCTGCAAAACGAAGGTTGCCAGGAGCCGTCT 2090
QY 1921 TTTTGGTGTGCTGATTTATCTAATCTTTCATTAAGATATGATACAAAACAGACGCGG 1980
Db 2091 TTTTGGTGTGCTGATTTATCTAATCTTTCATTAAGATATGATACAAAACAGACGCGG 2150
QY 1981 TTTGCGCATTTGAGTGGCGGTTATGTCATAGAGAGAAACCTACATCTTGTTCAGATATAG 2040
Db 2151 TTTGCGCATTTGAGTGGCGGTTATGTCATAGAGAGAAACCTACATCTTGTTCAGATATAG 2210
QY 2041 ATTCTACTGCTGATTTTGTACAGCTCTTGAAGAGATATAGAGTACTTGTGTAGCTAG 2100
Db 2211 ATTCTACTGCTGATTTTGTACAGCTCTTGAAGAGATATAGAGTACTTGTGTAGCTAG 2270
QY 2101 AATCAAGGTATACAGTCTACGGAGGAAGCTCTATTACAGCAACAGAAACCTATATCTCT 2160
Db 2271 AATCAAGGTATACAGTCTACGGAGGAAGCTCTATTACAGCAACAGAAACCTATATCTCT 2330
QY 2161 CTTCCTTGCAAACTACAGGCGCTTGTGCTTATGTTCTTACAGAGATTCCTGTTCTC 2220
Db 2331 CTTCCTTGCAAACTACAGGCGCTTGTGCTTATGTTCTTACAGAGATTCCTGTTCTC 2390
QY 2221 TTTTCAGAAACCTTAGCTACACCATACGATACGATACGATACGATACGATACGATACGATAC 2280
Db 2391 TTTTCAGAAACCTTAGCTACACCATACGATACGATACGATACGATACGATACGATACGATAC 2450
QY 2281 TATCTACTGTTTAAAGGAAGCTGGGGGAATGATAGTTTCGCTTTAGAAATTCGGTGAAGA 2340
Db 2451 TATCTACTGTTTAAAGGAAGCTGGGGGAATGATAGTTTCGCTTTAGAAATTCGGTGAAGA 2510
QY 2341 GCTCGATTTGCTTAGATGAAGAGTGTCTATTTGAGCAGTACATGCTCCCTCATGAATTTG 2400
Db 2511 GCTCGATTTGCTTAGATGAAGAGTGTCTATTTGAGCAGTACATGCTCCCTCATGAATTTG 2570
QY 2401 CAGTTTGTCTATGCAATCAGAGAGGTTTAAAGAACAGGGAACAGAAAGCTGCGGAATTT 2460
Db 2571 CAGTTTGTCTATGCAATCAGAGAGGTTTAAAGAACAGGGAACAGAAAGCTGCGGAATTT 2630
QY 2461 GGAAGTACCGCTCTTGTGAATCTTGCTTACCTATCGGAGTCCGATTTGATTAAGAAATCA 2520
Db 2631 GGAAGTACCGCTCTTGTGAATCTTGCTTACCTATCGGAGTCCGATTTGATTAAGAAATCA 2690
QY 2521 GACTGCCAAGATGACGATACAAATCTTAATCTTGTATATGATGAGATCTTGTCTGATG 2580
Db 2691 GACTGCCAAGATGACGATACAAATCTTGTATATGATGAGATCTTGTCTGATG 2750
```

gene 615..2957  
/gene="bnp\_3.2"  
/note="CPn0015"  
3196..4971  
/gene="bnp\_4.1"  
/note="CPn0016"  
3196..5953  
/note="Polymorphic Outer Membrane Protein G Family:  
contains frame shift"  
4874..5953  
/gene="bnp\_4.2"  
/note="CPn0017"  
6302..7786  
/gene="bnp\_5.1"  
/note="CPn0018"  
6302..9139  
/note="Polymorphic Outer Membrane Protein G Family:  
contains frame shift"  
7787..9139  
/gene="bnp\_5.2"  
/note="CPn0019"  
complement(9380..11467)  
/gene="CPn0020"  
complement(9380..11467)  
/gene="CPn0020"  
/codon\_start=1  
/transl\_table=1  
/product="Predicted OMP [leader peptide: outer membrane]"  
/protein\_id="AAD18173.1"  
/db\_xref="GI:4376272"  
/translation="MKRCFLFASFLVLMGSSADALTHQEAQKKNLYSHKSVGIV  
TIEGVLAHNNLRIOAKVYVENVGSLKIVAGNVMVYRAKTIQCDVLEYEDT  
DSCLTJNGRFAMYPWFLGSMITLPERIVIRKGYISREGPKDCLSGDLEYEDT  
SLISGKTLKRCRPIELFEPFSIMPEIRKPIFNFGSGKGLSGVLSYPSIR  
KHFSSTPELSEFNGHGVGNLHCSOKVPEENVPMKSYAHLAIDMAAHRBYRL  
HGDCFTKHKYNSGEYHLSDSMEVADIFPNNNKATGPTPDCITNDYFESYTL  
SSVANSFQANQELPYTLKQYPLISYNTVLENIYECGYLNFADFHDIVGEFSS  
LRLAARPIAKHTVPLIGTLLSTLSSSLIYSDVPEISRSOSLAKIQLDYRFLHK  
SYIORKHIEFVETIETRPDLAKNEDYIFSIDAFSLMLKAGIDTVSLKTNPR  
PPIHAKLMTHTHLSNTEKPLPPTACELSLPGKMTVSLDAEMWIKKHCWDMNI  
RWENIGNVAMTLESLSHRSKYSLKCDRENFIIIDVSRPIQLLSDPLSDNRILGK  
LFVPHPCWNYRLSRIGWHRQDTPNYETOMILGTLFEMQLKGYTEREADSRFF  
FFLKLDRKPP"  
complement(11484..13190)  
/gene="CPn0021"  
complement(11484..13190)  
/gene="CPn0021"  
/codon\_start=1  
/transl\_table=1  
/product="Predicted OMP [leader peptide]"  
/protein\_id="AAD18174.1"  
/db\_xref="GI:4376273"  
/translation="MGLFHLTLFGILLCSLPISIVAKPESVGHILYISNOQAL  
ATYLEADYVDHDFVIRKIGEDYIKOSIHSSPQTRKSTITIGLAGSSEADVLS  
QAMETADFLQQLVLSAVSGHLGKTSDDLKALASPVPVRLLEAVYLANUKTVL  
DHLHSFHLKPEEIQCSAALFLRIEESDAYIRDLAAKSAIRSATLQIIEYQ  
KRFLPRLNLTKSAPQDEALILYALGKLGOSYVINKLOQPDVUTVLAQAVALI  
ALGREEDALPVIRKQALEERPRALYHLHLSSEIGYIPALPIFLTKKSEKLVNVALI  
LLEGGDPLKLEYITERLVQPHYNETALHLSSEIGYIPALPIFLTKKSEKLVNVALI  
LSTRGEOITLFLRPLKAYLPCIKILASOKTOATRAISPLSTSHQEDLL  
FOAAKLPEEPIIRAYADAIYNLTRDKPKKSLHDYAKLLQETLLFVDTNQRHPS  
MPYLYOVTPESTRKMLDILETLATSSSDIRILLIQLMTEGDAKNPVLALGLIKI  
VE"  
complement(13172..13762)  
/gene="maf"  
/note="CPn0022"  
complement(13172..13762)  
/gene="maf"  
/codon\_start=1  
/transl\_table=1  
/product="Maf-type protein"  
/protein\_id="AAD18175.1"  
/db\_xref="GI:4376274"

/translation="MSLPVLVGSPPRRKFLIEKRPVTVIPSNFDESKVSGDPI  
AYQELAAQAVAVSELHSPDCILITGDTIVSYGRIFETKPODADAIQMLKTLRQ  
THDYVTSIAVLHKGKILFGSETSOILSMIDPHRIEESYIDYVGLNNGAVDCHGGL  
ILKRVHGVVVOGLPIOTLKYLEELNIDMDYSI"  
complement(13791..15383)  
/gene="yjkK"  
/note="CPn0023"  
complement(13791..15383)  
/gene="yjkK"  
/codon\_start=1  
/transl\_table=1  
/product="ABC Transporter Protein Appase"  
/protein\_id="AAD18176.1"  
/db\_xref="GI:4376275"  
/translation="MSIYLDKIGSLGRILFDDVSVYFNPNGCYGLNPGAGKSTL  
LKIINGMEPRGSLSPKYGILLRQNDSDHDTLQVINGNRLNENLQORNL  
LQEPDAILGMELEIEITIGBNGTRADEBELTGLIGIPREMIDKMAPIIDLOF  
RVLLQALFGHEALDLDEPNHLDLISINMGNFLKDYEGTVIYVSHDRFLNTIT  
HIADIDYDTIIIPNGYDMEKTAKEOKADIKSEKISQLEKFAKAGASRA  
SOVOSRIEIKKLOPELKSNIORPYIRPLSDSKSVLSLEAITKRGDHOVIH  
PSLEIYOGDKLGIIGNGIKTTLMLKLAGVAPSSGSLIGHOAISCYEPONSDV  
LADGQETLPEMIRRKGTINDQETRSYLGKMLRGDPAFQIOALSGETARILMAG  
MLEENHNYLIDBANNHLDLSVSALSMALNDYKTAIFVSHDKLIDPCATKLLIFD  
KDKITFEDGTVDYTAGHKQL"  
complement(15438..16376)  
/gene="CPn0024"  
complement(15438..16376)  
/gene="xerc"  
/codon\_start=1  
/transl\_table=1  
/product="Integrase/recombinase"  
/protein\_id="AAD18177.1"  
/db\_xref="GI:4376276"  
/translation="MASISFIDLVKVSASPHLLNVCOLDNLKIFLEBGNLA  
PSSPQDLTEKRYKSELPFSLETKHVMYIAKLLENKAKRTIKSLSSIKSRHYC  
VIOKLLENPAETLHGPRLKELPSMPYAOVEYLMAPDLSKYLGRDCLMEFYS  
SGLRSEIYVANKQDFDLSHLIRIRNGKERRIIPVSNAIOWIOIYLNHPRKRE  
KDPQAFINFRGRRISTRSIDRSFOEYIRSGLSGHIIPHTIRHTIATHWESGMDLK  
TIOALHGSSETTYVYQVSVKLRKHQHOEHNHRA"  
BASE COUNT 4710 a 3330 c 3665 g 4743 t  
ORIGIN  
Query Match 99.5%; Score 2773.4; DB 1; Length 16448;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2785; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
OY 1 ATGAAGCTCTTCCCTCCCAAGTTGTATTTCTACATTTGCTATTTCCCTTGCTATG 60  
|||||  
DB 172 ATGAAGCTCTTCCCTCCCAAGTTGTATTTCTACATTTGCTATTTCCCTTGCTATG 231  
OY 61 ATTGCTACCGAAGCTTTTGATTCAGTGGAGATTGGAGGAATAAATGCTAT 120  
|||||  
DB 232 ATTGCTACCGAAGCTTTTGATTCAGTGGAGATTGGAGGAATAAATGCTAT 291  
|||||  
OY 121 TTTTGTAGTCGTAAGAGTCAGGAAGTGTGGAAGTCTACTACTATTTAAAGGAATGTC 180  
|||||  
DB 292 TTTTGTAGTCGTAAGAGTGTGGAAGTGTGGAAGTCTACTACTATTTAAAGGAATGTC 351  
|||||  
OY 181 ACTGTAGAAAATATTTCTGTGGAACAGGACAGCAATTCACAAAAGCTGTTTAAACAAC 240  
|||||  
DB 352 ACTGTAGAAAATATTTCTGTGGAACAGGACAGCAATTCACAAAAGCTGTTTAAACAAC 411  
|||||  
OY 241 AAGGCGATTTACCTTTCACAGTAACGGAAGCTCTATTGTTCCAAAGGTTGATGCA 300  
|||||  
DB 412 AAGGCGATTTACCTTTCACAGTAACGGAAGCTCTATTGTTCCAAAGGTTGATGCA 471  
|||||  
OY 301 GGGACTGTAGCAGGGGCTGCTTAAACAGCAGCTGTGATTAATCTACACGTTTATA 360  
|||||  
DB 472 GGGACTGTAGCAGGGGCTGCTTAAACAGCAGCTGTGATTAATCTACACGTTTATA 531  
|||||  
OY 361 GGGTTTCTTCCGCTATCTTTTATTTGCGCTCTCTGGAAGTGTGATTAACGGCAAGGA 420  
|||||



```

|||||
Db 1141 GGTGAGCGCATCTACTTCTATGATCCATCTACTAGAGATTTCCACAGAGTATACAT 1200
|||
Qy 1201 GCTTTAAAGTAAATGAGACTCCGAGATTTCTGCACTACATATACAGGAGATCATC 1260
|||
Db 1201 GCTTTAAAGTAAATGAGACTCCGAGATTTCTGCACTACATATACAGGAGATCATC 1260
|||
Qy 1261 TTCCAGAGAGAAAGTATCGAGACAGAGCGGAGATTTCTAAAGATTTCTAGTTCGAG 1320
|||
Db 1261 TTCCAGAGAGAAAGTATCGAGACAGAGCGGAGATTTCTAAAGATTTCTAGTTCGAG 1320
|||
Qy 1321 CTACTAGAGCGCTGATCTCTTTCAGAGAGTCTATCTTTTAAACATGAGTACTCTG 1380
|||
Db 1321 CTACTAGAGCGCTGATCTCTTTCAGAGAGTCTATCTTTTAAACATGAGTACTCTG 1380
|||
Qy 1381 CAGACTAGAGCATCTACTCAACAGAGAGATTTCTGCTCGAATGAGAGTATGAGTACT 1440
|||
Db 1381 CAGACTAGAGCATCTACTCAACAGAGAGATTTCTGCTCGAATGAGAGTATGAGTACT 1440
|||
Qy 1441 CTAGAACCTGCTGATCTAGAGCAGCATTAACAAATTTGGTCATTAATCAGTCTATGAG 1500
|||
Db 1441 CTAGAACCTGCTGATCTAGAGCAGCATTAACAAATTTGGTCATTAATCAGTCTATGAG 1500
|||
Qy 1501 GGTGCAAGAGAGCAAAATAGAAACCAAGCTACGTCAAAAATCTGACTTTATCTGGA 1560
|||
Db 1501 GGTGCAAGAGAGCAAAATAGAAACCAAGCTACGTCAAAAATCTGACTTTATCTGGA 1560
|||
Qy 1561 ACCATCTATTATTTGAGACCGAGCGGACGTTTATGAAATCATAGTTTAAAGAAATCT 1620
|||
Db 1561 ACCATCTATTATTTGAGACCGAGCGGACGTTTATGAAATCATAGTTTAAAGAAATCT 1620
|||
Qy 1621 CAGATCTAGAGCATCTTAGAGCTCAAGAGCTCTGAACTGTAAAGAGACCGAGTACT 1680
|||
Db 1621 CAGATCTAGAGCATCTTAGAGCTCAAGAGCTCTGAACTGTAAAGAGACCGAGTACT 1680
|||
Qy 1681 CCAGATCTCTATTAATGAGTGAAGAAATTCATATAGAGTATGAGTATGAGTATGAG 1740
|||
Db 1681 CCAGATCTCTATTAATGAGTGAAGAAATTCATATAGAGTATGAGTATGAGTATGAG 1740
|||
Qy 1741 GTTTGGGGGAGAGGGGCTTCTACAGCTGCAACCTTCAACTGAGTAAAGTCTGATAT 1800
|||
Db 1741 GTTTGGGGGAGAGGGGCTTCTACAGCTGCAACCTTCAACTGAGTAAAGTCTGATAT 1800
|||
Qy 1801 CCTATATCCCGAGCTATGAGGCTTTTATGCTTATAGCTTATGAGTATGAGTATGAG 1860
|||
Db 1801 CCTATATCCCGAGCTATGAGGCTTTTATGCTTATAGCTTATGAGTATGAGTATGAG 1860
|||
Qy 1861 ATTAGCTCTCTCATATATCTTATGAGAGACTGCAAGAGGTTGCAAGGAGACCTGCT 1920
|||
Db 1861 ATTAGCTCTCTCATATATCTTATGAGAGACTGCAAGAGGTTGCAAGGAGACCTGCT 1920
|||
Qy 1921 TTTTGGGTCGTGATATATCTTCTGCAATGAGTATGAGTATGAGTATGAGTATGAG 1980
|||
Db 1921 TTTTGGGTCGTGATATATCTTCTGCAATGAGTATGAGTATGAGTATGAGTATGAG 1980
|||
Qy 1981 TTTTGGGTCGTGATATATCTTCTGCAATGAGTATGAGTATGAGTATGAGTATGAG 2040
|||
Db 1981 TTTTGGGTCGTGATATATCTTCTGCAATGAGTATGAGTATGAGTATGAGTATGAG 2040
|||
Qy 2041 ATTTAGTCTGCTGATTTTGTACAGCTTTTGAAGAGATAGAGACTTTTGTAGTAA 2100
|||
Db 2041 ATTTAGTCTGCTGATTTTGTACAGCTTTTGAAGAGATAGAGACTTTTGTAGTAA 2100
|||
Qy 2101 AATCAGAGTACAGTCTACGAGAGAACTCTATATACAGCAACGAACCTATATCTCT 2160
|||
Db 2101 AATCAGAGTACAGTCTACGAGAGAACTCTATATACAGCAACGAACCTATATCTCT 2160
|||
Qy 2161 CTTCCTGGAACCTAGAGGCTTGTGCTGCTATATCTTCTATAGAGATTTCCGTTCTC 2220
|||
Db 2161 CTTCCTGGAACCTAGAGGCTTGTGCTGCTATATCTTCTATAGAGATTTCCGTTCTC 2220
|||
Qy 2221 TTTTCAGGAACCTTAGACTACACCATACGAGTAAAGATGAGAAACCAAGTATACACA 2280
|||

```

```

Db 2221 TTTTCAGGAACCTTAGACTACACCATACGAGTAAAGATGAGAAACCAAGTATACACA 2280
|||
Qy 2281 TATCTACTGTTTAAAGAGAGCTGGGGGAAATGATAGTTTCGTTTAAATTTGGTGGAGA 2340
|||
Db 2281 TATCTACTGTTTAAAGAGAGCTGGGGGAAATGATAGTTTCGTTTAAATTTGGTGGAGA 2340
|||
Qy 2341 GCTCCGATTTGCTTATGATGAAAGTCTCTATTTGAGCAGTACATGCCCTTATGAGAAAT 2400
|||
Db 2341 GCTCCGATTTGCTTATGATGAAAGTCTCTATTTGAGCAGTACATGCCCTTATGAGAAAT 2400
|||
Qy 2401 CAGTTTGTCTATGACATCAGAGAGTAAAGAAACAGAGAGAGAGTCTGATATTT 2460
|||
Db 2401 CAGTTTGTCTATGACATCAGAGAGTAAAGAAACAGAGAGAGAGTCTGATATTT 2460
|||
Qy 2461 GGAAGTACGCTCTGTTGAACTTGTCTTACTTATGCGGATCCGATTTGATAGAGAA 2520
|||
Db 2461 GGAAGTACGCTCTGTTGAACTTGTCTTACTTATGCGGATCCGATTTGATAGAGAA 2520
|||
Qy 2521 GACTGCCAAGATGACATCAATCTGTTGTTATCTGATCTTGTCTAGT 2580
|||
Db 2521 GACTGCCAAGATGACATCAATCTGTTGTTATCTGATCTTGTCTAGT 2580
|||
Qy 2581 AACCOCGACTGTACGACAACTGCGAATTAAGCGTGTCTTGGAAAACTTGGTACG 2640
|||
Db 2581 AACCOCGACTGTACGACAACTGCGAATTAAGCGTGTCTTGGAAAACTTGGTACG 2640
|||
Qy 2641 AATTTGGCAACAAAGCTTATGCTCTGTCGAGGAAACATTTTGTCTTAACTCAAT 2700
|||
Db 2641 AATTTGGCAACAAAGCTTATGCTCTGTCGAGGAAACATTTTGTCTTAACTCAAT 2700
|||
Qy 2701 TTTGAGGCTTTAGCAATTTCTTTTGAATTTGGTGGTCTATCGCAATTAATGTA 2760
|||
Db 2701 TTTGAGGCTTTAGCAATTTCTTTTGAATTTGGTGGTCTATCGCAATTAATGTA 2760
|||
Qy 2761 GACTTAGAGCAAAATACCAATCTAA 2787
|||
Db 2761 GACTTAGAGCAAAATACCAATCTAA 2787
|||

```

**RESULT 2**  
**LOCUS** AE001587 16448 bp DNA BCT 01-DEC-2000  
**DEFINITION** Chlamydia pneumoniae section 3 of 103 of the complete genome.  
**ACCESSION** AE001587 AE001563  
**VERSION** AE001587.1 GI:4376271  
**KEYWORDS**  
**SOURCE** Chlamydia pneumoniae CWL029.  
**ORGANISM** Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
**REFERENCE**  
**AUTHORS** Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,  
Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.,  
Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
Nat. Genet. 21 (4), 385-389 (1999)  
**MEDLINE** 99206606  
**PUBMED** 10192388  
**REFERENCE** 2 (bases 1 to 16448)  
**AUTHORS** Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,  
Grimwood, J., Davis, R.W. and Stephens, R.S.,  
Direct Submission  
Submitted (01-DEC-1998) Program in Infectious Diseases, University  
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
**FEATURES**  
**source**  
1..16448  
/organism="Chlamydia pneumoniae CWL029"  
/strain="CWL029"  
/db\_xref="taxon:115713"  
172..2957  
/misc\_feature  
/note="Polymorphic Outer Membrane Protein G Family;  
contains frame shift"  
172..705  
gene  
/gene="pmp\_3.1"  
/note="Cpn0014"



c	9	536	19.2	10757	1	AE001628
c	10	535	19.2	2815	2	AE01829
c	11	535	19.2	17280	2	CP133034
c	12	528.6	19.0	2757	9	AE01837
c	13	517.2	18.6	10026	1	AE002193
c	14	517.2	18.6	300650	2	AE002546
c	15	517	18.6	2787	9	AE01839
c	16	517	18.6	10757	1	AE001628
c	17	517	18.6	17280	2	CP133034
c	18	515.4	18.5	12676	1	AE002192
c	19	514	18.4	2793	9	AE01841
c	20	514	18.4	15068	1	AE001627
c	21	510.8	18.3	3000	9	AE01851
c	22	510.8	18.3	6030	2	CP0MP54
c	23	465.8	16.7	3200	9	AE01827
c	24	465.8	16.7	6030	2	CP0MP54
c	25	380	13.6	2838	9	AE01849
c	26	346.8	12.4	6110	2	CP065942
c	27	346.8	12.4	6234	2	CP065943
c	28	343.4	12.3	10574	1	AE001586
c	29	338.6	12.1	2526	9	AE01833
c	30	271	9.7	10044	1	AE002237
c	31	270.2	9.7	10766	1	AE001585
c	32	270	9.7	3052	9	AE01831
c	33	244.8	8.8	10695	1	AE001629
c	34	244.8	8.8	15789	1	AE002191
c	35	207.6	7.4	1830	9	AE01855
c	36	197.6	7.1	840	9	AE01843
c	37	135.4	4.9	14177	1	AE001360
c	38	127.6	4.6	14482	1	AE002293
c	39	104	3.7	1200	9	AE01853
c	40	80.8	2.9	1545	9	AE01845
c	41	68	2.4	216	2	CP059306
c	42	61.2	2.2	10001	1	AE002248
c	43	61.2	2.2	11781	1	AE001676
c	44	61.2	2.2	325865	2	AE002548
c	45	54.2	1.9	10908	1	AE002294

ALIGNMENTS

RESULT	1	2787 bp	DNA	PAT	21-JAN-2000
LOCUS	AB1835	Sequence 9 from Patent WO9858953.			
DEFINITION	AB1835				
ACCESSION	AB1835.1	GI:6731868			
VERSION					
KEYWORDS					
SOURCE		unidentified.			
ORGANISM		unidentified.			
REFERENCE		1 (bases 1 to 2787)			
AUTHORS		Madsen, A. and BirkeLund, S.			
TITLE		NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE			
JOURNAL		Patent: WO 9858953-A 9 30-DEC-1998;			
FEATURES		MADSEN ANNA SOFIE (DK); BIRKE LUND SYVEND (DK)			
source		Location/Qualifiers			
		1..2787			
		/organism="unidentified"			
		/db_xref="taxon:32644"			

BASE COUNT 811 a 583 c 598 g 795 t

Query Match 100.0%; Score 2787; DB 9; Length 2787;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2787; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAAGCTCTTCCCAAGTTGATTTCTACATTTGCTATTTCCCTTGCTATG 60  
1 ATGAAGCTCTTCCCAAGTTGATTTCTACATTTGCTATTTCCCTTGCTATG 60

QY	61	ATTGCTACCGACAGAGTTTGGATTCAAGTCGAGATTCGATGGGAATAAATGTAAT	120
DB	61	ATTGCTACCGACAGAGTTTGGATTCAAGTCGAGATTCGATGGGAATAAATGTAAT	120
QY	121	TTTTTCAGTTTCGAGAGTATGAGAAAGTGTGGAATCTACTATTTAAAGGAATGTC	180
DB	121	TTTTTCAGTTTCGAGAGTATGAGAAAGTGTGGAATCTACTATTTAAAGGAATGTC	180
QY	181	ACTCTAGAAAATATTCCTGGAACAGCAGCAATCAAAAAGCTGTTTAAACACT	240
DB	181	ACTCTAGAAAATATTCCTGGAACAGCAGCAATCAAAAAGCTGTTTAAACACT	240
QY	241	AAGGCGATTTGACTTTCACAGTAACGGAACTCTATTTGTCACAAAGGTGATGCA	300
DB	241	AAGGCGATTTGACTTTCACAGTAACGGAACTCTATTTGTCACAAAGGTGATGCA	300
QY	301	GGGACTGTAGCAGGGGCTGCTTTAACACAGCGTGTGATTAATCTACACGTTTATA	360
DB	301	GGGACTGTAGCAGGGGCTGCTTTAACACAGCGTGTGATTAATCTACACGTTTATA	360
QY	361	GGGTTTCTGCTATCTTTATTTGCGTCTCGGAAGTTCATACCTACCGCAAGGA	420
DB	361	GGGTTTCTGCTATCTTTATTTGCGTCTCGGAAGTTCATACCTACCGCAAGGA	420
QY	421	GCCGTAGCTGCTTACGGGTAGCTTGAAGTTGACAAAATGTACGTTGCTTACAC	480
DB	421	GCCGTAGCTGCTTACGGGTAGCTTGAAGTTGACAAAATGTACGTTGCTTACAC	480
QY	481	AAAACTTTTCAACGATATATGGCGGTGTATCACCGCAAAAACCTTTCATTAACAGG	540
DB	481	AAAACTTTTCAACGATATATGGCGGTGTATCACCGCAAAAACCTTTCATTAACAGG	540
QY	541	ACTCAATGTACGCTCTGTTTCTGAAAATACCTCCTCAAAAGAGCGGAGCCATTGAG	600
DB	541	ACTCAATGTACGCTCTGTTTCTGAAAATACCTCCTCAAAAGAGCGGAGCCATTGAG	600
QY	601	ACTTCGATGCCCCCTTACCAATTAAGTGAACCAAGGGGAAGTCTTTTTCGACAATCT	660
DB	601	ACTTCGATGCCCCCTTACCAATTAAGTGAACCAAGGGGAAGTCTTTTTCGACAATCT	660
QY	661	TCTTGGAATTCGGAGCTGCATTTTACAGAAGCTCGGTACATTTCTAATATGCT	720
DB	661	TCTTGGAATTCGGAGCTGCATTTTACAGAAGCTCGGTACATTTCTAATATGCT	720
QY	721	AAAGTTTCTTATTCACAAATTAAGTCAAGAGCGAGCTCCTCAACACGGGGATATG	780
DB	721	AAAGTTTCTTATTCACAAATTAAGTCAAGAGCGAGCTCCTCAACACGGGGATATG	780
QY	781	TCAGAGAGTGTATCTGTGCTTATAAACTAGTACAGATCTAAGGTCACTCACTGGA	840
DB	781	TCAGAGAGTGTATCTGTGCTTATAAACTAGTACAGATCTAAGGTCACTCACTGGA	840
QY	841	AATCAGATTTACTCTTCAGCAACAATATATGCAACACGGGGAGACTATCTATGTC	900
DB	841	AATCAGATTTACTCTTCAGCAACAATATATGCAACACGGGGAGACTATCTATGTC	900
QY	901	AAAAAGCTGCACTGCTCCGAGAGCTTACCTTTAGTAGAATAATGTGTCAATGGA	960
DB	901	AAAAAGCTGCACTGCTCCGAGAGCTTACCTTTAGTAGAATAATGTGTCAATGGA	960
QY	961	GCTACAGCTCTAAAGGTGAGCCATAGTATCGAAGATAGTGGGAATGAGTTATCC	1020
DB	961	GCTACAGCTCTAAAGGTGAGCCATAGTATCGAAGATAGTGGGAATGAGTTATCC	1020
QY	1021	GCGATAGTGTGACATTTCTTTTAAAGGATACAGTCTATCTCTCTGAGAG	1080
DB	1021	GCGATAGTGTGACATTTCTTTTAAAGGATACAGTCTATCTCTCTGAGAG	1080
QY	1081	AATTAAGATGATGAGCTTACGAAAGAGTGAAGATGACAGTTTGGTCTGCTGCT	1140
DB	1081	AATTAAGATGATGAGCTTACGAAAGAGTGAAGATGACAGTTTGGTCTGCTGCTGCT	1140
QY	1141	GGTAGAGCCATCTACTTATATATCCATTAATACAGATCTTCCACAAAGTTACAGAT	1200

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 01:09:34 ; Search time 18674.8 Seconds

(without alignments)  
2308.379 Million cell updates/sec

Title: US-09-446-677B-9

Perfect score: 2787

Sequence: 1 ATGAGGTCTTCTTCCCA.....GAGCAAAATACCAATTCNA 2787

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_ba3: \*  
4: gb\_in1: \*  
5: gb\_in2: \*  
6: gb\_in3: \*  
7: gb\_om: \*  
8: gb\_ov: \*  
9: gb\_pat1: \*  
10: gb\_pat2: \*  
11: gb\_ph: \*  
12: gb\_pl1: \*  
13: gb\_pl2: \*  
14: gb\_pl3: \*  
15: gb\_pl4: \*  
16: em\_ba1: \*  
17: em\_ba2: \*  
18: em\_fun: \*  
19: em\_htgo\_hum: \*  
20: em\_htgo\_inv: \*  
21: em\_htgo\_rnd: \*  
22: em\_htg\_hum1: \*  
23: em\_htg\_hum2: \*  
24: em\_htg\_hum3: \*  
25: em\_htg\_hum4: \*  
26: em\_htg\_hum5: \*  
27: em\_htg\_hum6: \*  
28: em\_htg\_hum7: \*  
29: em\_htg\_hum8: \*  
30: em\_htg\_inv1: \*  
31: em\_htg\_inv2: \*  
32: em\_htg\_other: \*  
33: em\_htg\_rnd: \*  
34: em\_hum1: \*  
35: em\_hum2: \*  
36: em\_hum3: \*  
37: em\_hum4: \*  
38: em\_hum5: \*  
39: em\_hum6: \*  
40: em\_hum7: \*  
41: em\_in: \*  
42: em\_om: \*  
43: em\_or: \*

44: em\_ov: \*  
45: em\_pat: \*  
46: em\_ph: \*  
47: em\_pl: \*  
48: em\_ro: \*  
49: em\_sts: \*  
50: em\_sy: \*  
51: em\_un: \*  
52: em\_vi: \*  
53: gb\_sts1: \*  
54: gb\_sts2: \*  
55: gb\_sts3: \*  
56: gb\_sy: \*  
57: gb\_un: \*  
58: gb\_v11: \*  
59: gb\_v12: \*  
60: gb\_htg1: \*  
61: gb\_htg2: \*  
62: gb\_htg3: \*  
63: gb\_htg4: \*  
64: gb\_htg5: \*  
65: gb\_htg6: \*  
66: gb\_htg7: \*  
67: gb\_htg8: \*  
68: gb\_htg9: \*  
69: gb\_htg10: \*  
70: gb\_htg11: \*  
71: gb\_htg12: \*  
72: gb\_htg13: \*  
73: gb\_htg14: \*  
74: gb\_htg15: \*  
75: gb\_htg16: \*  
76: gb\_htg17: \*  
77: gb\_htg18: \*  
78: gb\_htg19: \*  
79: gb\_htg20: \*  
80: gb\_htg21: \*  
81: gb\_htg22: \*  
82: gb\_htg23: \*  
83: gb\_htg24: \*  
84: gb\_htg25: \*  
85: gb\_pr1: \*  
86: gb\_pr2: \*  
87: gb\_pr3: \*  
88: gb\_pr4: \*  
89: gb\_pr5: \*  
90: gb\_pr6: \*  
91: gb\_pr7: \*  
92: gb\_pr8: \*  
93: gb\_pr9: \*  
94: gb\_r01: \*  
95: gb\_r02: \*  
96: gb\_in4: \*  
97: gb\_pr10: \*  
98: em\_ba3: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	2787	100.0	2787 9 A81835	A81835 Sequence 9
2	2773.4	99.5	16448 1 AE001587	AE001587 Chlamydia
3	2773.4	99.5	26920 2 CPN133035	AT133035 Chlamydia
4	2771.8	99.5	12127 1 AE002235	AE002235 Chlamydia
5	2771.8	99.5	299650 2 AP002545	AP002545 Chlamydia
6	548.2	19.7	4926 2 CP072499	U72499 Chlamydia
7	536.6	19.3	12676 1 AE002192	AE002192 Chlamydia
8	536.6	19.3	300650 2 AP002546	AP002546 Chlamydia

**BEST AVAILABLE COPY**

---

```
Db 1935 tccctacagtagcaactcctcactagcactatggtatcaagtgactcttgggaatgacttgggt 1994
QY 1746 GG-----GGACAGGGGGCTTCTACGACTGCAACCTTCAACTGACTAAACTGG 1793
Db 1995 tgatgataccgcaagcactccaagaactaagacgacgacattagcttggaccaatcacg 2054
QY 1794 CTATATTGCTTAATCCCGAGCTATGCGGCTTTTGTCCCTAATAGCTTATGGAATGCATT 1853
Db 2055 ctacctccgaatcccgagcgcaagaccccttctcctaataagcccttggggatcctt 2114
QY 1854 TATAGATATTAGCTCTCTCCATTACTTATGAGACTGCAACGAAAGGGTTGCAAGGAGA 1913
Db 2115 ttcaagaatccaagcgatccaagtgatagagagaagtgcttggactcttggtaaga 2174
QY 1914 CCGTGCCTTTTGGTGTGCTGATATCTTCTTCATTAAGATAGTACAAAACACG 1973
Db 2175 tcgaggtctctgggtcgtcgggagtcgcacattctctagataagaataagaaaggaa 2234
QY 1974 ACGGGGGTTTGGCATTTGAGTGGCGGTTATGTCATAGAGAAACCTACATACTTGTTC 2033
Db 2235 acgcaaatccgcatcaaatctgltgltgaltgcatcggaagtcgaagcgaacttggtc 2294
QY 2034 AGATAGATTTCTTATGCTGATTTTGTGCACTCTTTGGAAGAGATAGACTACTTGT 2093
Db 2295 tgaataacttaattagcttggcttggccaactcttggtagcgataaagaattccttagt 2354
QY 2094 AGCTAAGATCAGGTACAGTCTACGAGAGAACTCTCTATTACGACACAAGAAACCTA 2153
Db 2355 cgctaaaaatcatactgatactatgacgagccctctataccaacacattacagaa 2414
QY 2154 TATCTCTCTTCTGCAACCTACGCGCTTGTCTGTATGTTCTTACAGAGATTTC 2213
Db 2415 tagtgggtctcactaggttgc---ctagataaaacttccgtctcttggagtcataaac 2471
QY 2214 TGTCTCTTTTGGAGAAACCTTACCTACACCATACGATACGATACGATGAAACCAAGTA 2273
Db 2472 cctcgttcttgaagggcagctcgcttataagcagtcagtaagtcgaaagaacaagta 2531
QY 2274 TACACATATTCCTACTGTTTAAAGAAAGCTGGGGAATGATAGTTTGGCTTGAATTGG 2333
Db 2532 taccggtatcctgagtggaaggttcttgggggaataatgctttaacatgaltgttgg 2591
QY 2334 TGGAAAGAGCCTCCGATTTGCTTAGATGAAAGTGCCTATTGAGCAGTACATGCCCTTCAT 2393
Db 2592 agcttcttctatcttactccttgaactcgtcaltgttltgatacctatgctccatcat 2651
QY 2394 GAAATTCAGTTTGTCTATGACATACAGAAAGTTTTAAAGACAGGAAACAGAACTCG 2453
Db 2652 caaactgaatctgacctatataagtcgagacaagcttccgagaagaagtlacagaagaag 2711
QY 2454 TGAATTTGGAAGTACCGCTCTTGTGAATCTTGCCTTACCTATCGGGATCCGATTTGATTA 2513
Db 2712 atcttctgatacagcaaccttccatcttcttgcctataaggggtgaagtttggaa 2771
QY 2514 GGAATCAGACTGCCAAGATGCAAGTACATCTAATCTTGTATACTGTGATCTTGT 2573
Db 2772 gtcttcgattgataagcttcttctatgatacttacttccatgttccctgactcat 2831
QY 2574 TCGTAGTAACCCGACGCTAGCAGACACATGCAATTAGCGGATTCCTTGGAAACCTT 2633
Db 2832 ccgcaatgatacccaatgcaactacagacactgtatacagcggaagcccttggaaactta 2891
QY 2634 CGGTACGAATTTGGCAAGACAAAGCTTTAGTCCCTTGTGACAGGAACATTTTGTCTTAA 2693
Db 2892 tgcataacttagcagcaagagccttgcgaagtgctgacagcgactacactacgctctc 2951
QY 2694 CTCAAATTTGAGCCTTAGCCCAATTTTCTTTGAATGCGTGGGTC 2741
Db 2952 tccatagtctgaagtgctcgccagttgtcttgaagttcgtgaltc 2999
```

Search completed: October 2, 2001, 06:01:31  
Job time: 60238 sec



OY	1041	CTTTTGGGAATCACTGACCTTCTTA---CTACTCCTGGGAGACATTAAGATGATGCA	1097
Db	1135	gtcccttggcaaaccttctaacccttaaccttcgcgcacacatcgcgaacggaatgcatcta	1194
OY	1098	CTTAGAACAAGTGCACAAAGATGACAGCTTGTGCTTCGTGGTAGAGCCACTACTT	1157
Db	1195	cttggagatcgttcagcaaaaatttaacgaacttaagggagcccaaggccaaatctcatattc	1254
OY	1158	CTATGATCCCATTAACCTACAGGATCTTCCACACAGTTACAGATGTCTTAAAGTTAATGA	1217
Db	1255	ctatgctccggt---tgcattcaaccacaaggagcttcagagcgttctcgcattcaacca	1311
OY	1218	GACTCCGGCAGATTTCTGCACCTACAAATATACAGGGAACATCATCTTCACAGAGAAATTT	1277
Db	1312	accggatagaacacctcgccttagattatcttaacgaacgatatttctctcggggaaaaagct	1371
OY	1278	ATTCAGAACAGAGGCGCCAGATCTCTTAATAATCTTACTTGGAAGCTACTACACCCGTATAC	1337
Db	1372	ctctgcacgataggaacggaagctcgtcgtataacttcacatctatataaagaacacctggc	1431
OY	1338	TCCTTTGAGGAGTACTCTATCTTTAAACATGAGAGTGACTCTGCAGACTCACGCAATTCAC	1397
Db	1432	cttagcccttcggaaccttagcactcaagaagaaatgcgagttagatgtcattatgtyttcac	1491
OY	1398	TCACAGACGAGATTTCTGCTGTCCGAATTTGACAGCTAGGAATACTCT---AGAACCTGCCTGA	1454
Db	1492	acaaactggaagpcccttaacctctctatgcacaacgaagaaacgaactcaagaacgatactga	1551
OY	1455	TATCAGCAGCCATTAACCAATTTGGTCTCATTTAATCATCATCTTATATAGACGGGTGCACAAAGAC	1514
Db	1552	agctatcagcttcttaaccaacttgcgttgatctcttctcgtcttaggggaaataaagaggt	1611
OY	1515	AAAAATGGAACCAAGCTACGTCAAAAATCTGACTTTATCTGCAACCAATCACTTATTT	1574
Db	1612	gtccattggaacacgacgagcgacaaacataactaactcaacctctctctgttttcca	1671
OY	1575	GGACCCGACGGGACGTTTATGAAATATATAGTTTAAATTTCTCAGTCTCCTACGACAT	1634
Db	1672	agataagtagcgcaatttcttaagaaacatcagaataaaccaagcttcaacgcacctt	1731
OY	1635	CTTAGACTCTAAAGCTCTTCTGGAAGT-----TAAACACACCGCAGTAGTCTCCAGA	1685
Db	1732	gggtgattatcacgctcgtcactcgtcgtcgtatcgatatttatctcgatgyccttctcaactc	1791
OY	1686	TCTATATATGGGTGAGAAATTCACATTACGGCTATTCAGGAGACTTG6GGCCCAATTGTTTG	1745
Db	1792	tccagtaacaacatccagaaacctcatcattacggtatccaggaacatcgggaagccacttggc	1851
OY	1746	GGGAGACGGGGCTTCTACGACTGCACACTTCAACTGTGACTAAATCGCTATATCTCTAA	1805
Db	1852	agacacatcaactcgcacaaatcagaagaaactcgtcgttgaactacagggctcaaacacctta	1911
OY	1806	TCCGACAGCATGGGCTTATAGTCCCTAATAGCTTAATAGCTTAATAGCATTTATTAAGATATTAG	1865
Db	1912	tccctgagcgtcagagcttcogtagttccgattcaattatgaggaatcccttactcgtcaactcg	1971
OY	1866	CTCTCTCCATTAATCTTAATGAGACTGCACAAACGAAGGTTGACGAGAGACCGTGCCTTTTGG	1925
Db	1972	caactctacagcagatcatcagcatctcaagcagatagtatctatcgaacgaagagactctg	2031
OY	1926	GTGTGCTGGAATTATCTAACTTCTTCCATAAGATATGACAAAACACACGCGGGTTTCG	1985
Db	2032	ggacatcaggaaactcgtgaatttcttccataaagataaataatcagaactaaccaagaactcgg	2091
OY	1986	CCATTTGAGTGGGGGTTATGTCATTAAGAGAGAAACCTACACTTGTTCAGATTAAGATTCCT	2045
Db	2092	acataaaagctacggtctatatgttctggaggaaggtcgtcgaagaatttcttcgaaaatacttc	2151
OY	2046	TAGTGTGCACTTTTGTCAAGCTCTTTTGAAGATATGAGACTCTTTGTAGCTAAGAATCA	2105
Db	2152	caagtgtagcttcttcgcagcgtctcgtgtaagaataaagaacctggtttatagttgaataaac	2211
OY	2106	AGGTACAGTCTACGAGAGACTCTCTATTACGAGACAACGAACCATTAATCTCTTCC	2165

Db	2212	ctctcataaccatttagcgcgtctataacctgcacactcgagcatcttcctcagagagctcc	2271
Qy	2166	TTGCAAACATACGAGCCCTTGTTCGTTGCTTATGTTCCTACAGAGATTCCTGTTCTCTTTTC	2235
Db	2272	--catgcccctcatttggaaagtatcaaccgcatctgcaagaagtattctccctcatttggaa	2328
Qy	2226	AGGAACCTTATAGCTACACCCATACGGATTAACGATCTGAAACCAACGATATACAACTATTC	2285
Db	2329	tgcacagctaacgctacacgctacataaataatgatacgtatacgcctatacttcctac	2388
Qy	2286	TACGTTTAAAGAGAGCTGGGGGAAATGATAGTTTCGGTTTAAAGATTCGGTGGGA--AGAGC	2342
Db	2389	tgaagcttaagctctctcgacccaataactctcggtgcctacaagctccgagagctctctg	2448
Qy	2343	TCCGATTTGCTTAGATGAAAGTGCTCTATTTTTCAGAGTACAGCCCTTCATGAAATTTGCA	2402
Db	2449	tctatatctccctaagaagaagcacgctcttcctcagagatacttcctcctcttaagttcca	2508
Qy	2403	GTTTGTCTATGACATACGAGAGGTTTAAAGAACAGGAGACAGAACTCGTGAATTTGG	2452
Db	2509	ggcagctctacagccgcgaacaaacttaagaagatggcgctgaagcccgctcttga	2568
Qy	2463	AAGTAGCGCTTGTTGAAATCTTGCTTACCTATCCGGAGATCCGATTTGATTAAGAAATCGA	2532
Db	2559	tgatgtagagacctagctgaactgcctctacccctgcgcgcatcttgataaataatcccgga	2628
Qy	2523	CTCGCAAGATGCAACGATCAATCTAACCTTTGGTTTACTGTGATCTGTTGCTGATGA	2582
Db	2629	agatgaaaaaataatttcgagattctctctagctacatctgtgtgctgatacgtaaaaa	2688
Qy	2583	CCCCGACTGTACGACCAACACTGCGGAATTAGCGGTATTTTGGAAAAACCTTCGATACAA	2642
Db	2689	tcccccgttcgctactctctctataatggctagtggaagcctcttggactctgcctatgaaaaa	2748
Qy	2643	TTTGGCAAGCAACCTTACTTACTCTGCTGGTGAGGAAACATTTTCTTAACTCAAAATTT	2702
Db	2749	ccctgcacagaaagcctctcttagcaagatgcggaagcactctgactctctccctcaatgt	2808
Qy	2703	TGAAGCTTTAGCCAAATTTTCTTTGAATTCGCTGGGTCATCTCGCAATTAACAATGTAGA	2762
Db	2809	agaactctctcgggaagctgcttaigactctggtgctcagaacacatcataatgtaga	2868
Qy	2763	CTTAGAGACAAATATACCAATTTCTA	2786
Db	2869	tttggggtctaagatactcatctca	2892
RESULT 15			
AAx06828 standard; DNA; 3000 BP.			
AC	AAx06828;		
DT	26-APR-1999	(first entry)	
De	Chlamydia pneumoniae surface exposed protein Omps DNA.		
KW	Omps: outer membrane protein 5; surface exposed protein; antigen;		
KW	infection; diagnosis; vaccine; atherosclerosis; asthma; ss.		
XX	Chlamydia pneumoniae.		
XX	OS		
XX	Key	Location/Qualifiers	
FT	CDS	259...3000	
FT		/*tag- a	
XX	W09858953-A2.		
XX	PN		
XX	PD	30-DEC-1998.	
XX	PF	19-JUN-1998.	98MO-DK00266.
XX	IX		

OY	2643	TTTGGCAGACAAAGCTTTAGCCCTTGTGAGGGAACCATTTTTCGTTTAAACCAATTT	2702
Db	2649	ccctgcacgaaaagcctctcttagcaaatgtctggaagccactctgactctccctccatgt	2708
OY	2703	TGAAGCCTTAGCCAAATTTCTTTTGAATTCGCTGGTCATCTCGCAATTTACAATGTAGA	2762
Db	2709	agaactctctgggaagctgctatgctctgcttggtcctcagcaacacatctacaatgtaga	2768
OY	2763	CTTAGGAGCAAAATGACCAATTCCTA	2786
Db	2769	ttgtggcctaagatactcattcta	2792
RESULT 14			
ID	AAA30853	standard; DNA; 3000 BP.	
XX	AAA30853:		
XX	29-AUG-2000	(first entry)	
DE	Chlamydia antigen CPN100639 full length coding sequence.		
KX	Chlamydia antigen; diagnosis; infection; community acquired pneumonia;		
KW	therapy; upper respiratory tract disease; bronchitis; sinusitis;		
KW	asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;		
XX	ds.		
OS	Chlamydia pneumoniae.		
FH	Key	Location/Qualifiers	
FT	CDS	101..2893	
FT		/*tag -a	
XX		/product= Chlamydia antigen CPN100639	
PN	MO200032794-A2.		
PD	08-JUN-2000.		
XX			
PF	01-DEC-1999;	99WO-CA01147.	
XX			
PR	01-DEC-1998;	98US-0110339.	
PR	01-DEC-1998;	98US-0110340.	
PR	01-DEC-1998;	98US-0110427.	
PR	01-DEC-1998;	98US-0110428.	
XX	01-DEC-1998;	98US-0110438.	
PA	(CONN-) CONNAUGHT LAB LTD.		
PI	Murdin AD, Oomen RP, Wang J;		
DR	WPI; 2000-412339/35.		
DR	P-PSDB; AAY90240.		
XX			
PT	Nucleic acids encoding polypeptide antigens from Chlamydia useful for		
PT	preventing, diagnosing and treating diseases such as community acquired		
PT	pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset		
PT	asthma -		
PS	Claim 2; Fig 7; 174pp; English.		
XX			
XX	This sequence encodes a Chlamydia antigen of the invention, designated		
CC	CNN100639. The nucleic acids (and their complementary sequences) may be		
CC	used as diagnostic agents for detecting the presence of nucleic acids		
CC	encoding Chlamydia antigens in samples according to standard methods,		
CC	and therefore, for diagnosing Chlamydia infections. For example, they may		
CC	be used as primers and probes for diagnostic polymerase chain reaction		
CC	(PCR) assays. Antisense sequences may be used to down regulate		
CC	expression of the proteins and may be used to treat infections. The		
CC	nucleic acids may also be used to produce the protein antigens they		
CC	encode according to standard recombinant DNA methodologies. The		
CC	proteins may then be used as antigens for the production of antibodies		
CC	(i.e. as vaccines) for preventing infection by Chlamydia. The		

[illegible]



Dd 453 tcttacccgataatggaacgatctctctcttagcccaaaacgtctccaaatgaagaactaataca 512  
Qy 501 TGGGGTGCTATACACCGAAAAACCTTTTCAATTACAGAGGACTACAAATGCTAGCTCTGT 560  
Dd 513 tggcggaagcagatcaccaaaaaaacctctctctatctctcgtggaatcctctctctcaactt 572  
Qy 561 TTTCGAAATACCTCTCTCAAGAAAGGCGAGCCATTACAGATTCGATGCCCTTACCAT 620  
Dd 573 cactcagtaatacgcgaaaaaaattagtggaagcgatcatatagctctcgtcgcaagatc 632  
Qy 621 TACTGGAACCAAGGGGAGTCCTCTTTTCTGACAAATCTCTTCGATTCGAGCTGC 680  
Dd 633 ttcagaaacacccgcagcttagctctctatgaataaagaagaatcgggagcggg 692  
Qy 681 AATTTTACAGAGCGCTGGTGACTATTCTTAATATGCTAAAGTTTCTTATGACAA 740  
Dd 693 tctgggtcttgaagccagctccctcgatctcaaaatagctccctctctctcggaa 752  
Qy 741 TTAAGGTACAGAGAGAGCTCTCAACAAAGGGGATATGTCAGAGGCTGCTATCTGTC 800  
Dd 753 cactgcaacgaatgctgagcgaag-----ggcggggccattatgtg 794  
Qy 801 TTATTAATCTAGTACAGATTAAGTCACTCCCTCACTGGAATCAGATGTTACTCTTCAG 860  
Dd 795 tgaaaaaacaggaagagactcctactctactactctctcgtgaaataaagctcgaactcgc 854  
Qy 861 CAACAATCATCGACAAAGAGGGGAGGAGTATCTATGTAAGAAAAAGCTCGAATCGCTTC 920  
Dd 855 cgaagaaccttcagtaactcaagcgagcgcaatctgcccactgctcagatctctcgc 914  
Qy 921 CGAGAGACTTACCTTATTCAGTAGAATAGTGTCAATGAGAGTACAGCTCTCAAAAGTGG 980  
Dd 915 tgcgtgaccttaacctattcaataatagatgcggaacacagctcgaagcaagggcg 974  
Qy 981 AGCCATACCTTTCGAAGTACTGGGGAATTGAGTTTATCCGCCCATGACTGATGATGTT 1040  
Dd 975 cgtatctgcaatccgcagctcgtctcttaagctctctcgtcgaataaagagagatcac 1034  
Qy 1041 CTTTTTAGGGAATACAGTCACTCTA---CTACCTCGGAGCAATGAGATGATGATGAG 1097  
Dd 1035 gtctcttggcaacacctcaacctcaacctcgcgcgaacatcgaaagaaatgcatctta 1094  
Qy 1098 CTTAGGAAGAGTGCMAAGATGACAGCTTTGCGTTCTGCTGCTAGAGCCATCTACTT 1157  
Dd 1095 ccttgagctgcgcagcaaaaattagaaacttaaggcagcccaagcgaatctatctatt 1154  
Qy 1158 CTATGATCCCAATACATACAGATCTTCCACAACAGTACAGATGCTTTAAAGTTATGA 1217  
Dd 1155 ctatgactcgat---tgcacttaaacacacagagagcttcgaacatcaacca 1211  
Qy 1218 GACTCCGCAATCTGTGCATACAAATATACAGGGAACATCATCTTCACAGAGAAAGTT 1277  
Dd 1212 acgagataagcaacctgcttagatctatcaggaacgactglatcttctgggaaagct 1271  
Qy 1278 ATCAGAGACAGAGCCGCGAGATTTCTAAATAATCTTACTCGAAGTACATACAGCCGTAA 1337  
Dd 1272 ctctcgagatgaagcgaaagctgctgataactcaactatataaagcaaccaatggc 1331  
Qy 1338 TCTTTCAAGAGGACTCTATCTTTTAAACATGAGTGACTCTGACAGCTCAGGCAATTGAC 1397  
Dd 1332 tctagccctctggaaccttagcactcaagaagaatgctgaagtagtgcacatggttcac 1391  
Qy 1398 TCACAGAGCAATCTCTCTCGAATGAGAGTGAAGTACTCT---AGAACCTGCTGA 1454  
Dd 1392 acagactgaagcctctcaactctcaatgcaacaggaacaaagctcaagagatacga 1451  
Qy 1455 TACTAGACACCTTAAACATTTGGTCATTAATCATGATGTTCTATAGAGCGTGCAGAAAGC 1514  
Dd 1452 agcatcaagctctcaaaactgctgactcttctcgtccttagagggaaataaagagct 1511  
Qy 1515 AAAAATGAAACCAAGCTACGTAAGAAAAATCTGACTTTATCTGGAACCAATCATTAT 1574  
Dd 1512 gtccattgaaacagcaggaagcacaacataacttaacctctactctctgtttcca 1571

Qy 1575 GGACCCGAGGGGACAGTATTATGAATAATCATATGTTTAAAGAAATCCAGTCCATGACAT 1634  
Dd 1572 agatagtagcggcaacttttatgaaagccaatcagtaaaccaagctcacaagccttt 1631  
Qy 1635 CTTAGAGCTCAAGGCTCTGTGAACTG-----TAACAAAGCAGCGAGTACTCAGA 1685  
Dd 1632 ggtgtatcactgctgctactgctcagcgatattatctcgtatcgtctctcacttc 1691  
Qy 1686 TCTTATATGGGTGAGAAATTCATTAAGGCTATAGGGAATCTGGGCCCAATGTTTG 1745  
Dd 1692 tccagtaacaaactccaagaacctcaatcagggatcaagggaacttgggaagccaattggc 1751  
Qy 1746 GGGAGAGGGGCTTCTACAGCTGCAACCTTCAACTGAGTAAATCGACTATATCTTAA 1805  
Dd 1752 agacacatcaactgcaaaatacaggaactatgacttgggtaactaagggctacaacctaa 1811  
Qy 1806 TCCGAGGGTATCGGCTCTTATGATCCCTAATAGCTTATGGAATGCAATTAATGATATAG 1865  
Dd 1812 tccgtgagctagagcttcgtagtcccgatctcatattatgggcatctctactgacatcgc 1871  
Qy 1866 CTCTCTCCATTAATCTTATGAGAGCTGCAACGAGGTTGCCAGGAGACCGTCTTTTGG 1925  
Dd 1872 cacttcagcagatcagatcagacatcctcaagcgaatagatctactcaagcagagagctctg 1931  
Qy 1926 GTGTGCTGATTAATCACTTCTTCCATAGAGATGTAACAAAACACGAGCGGGGTTTCG 1985  
Dd 1932 ggcatacgaagaactgcgaactctctccataaagataaatacgaagaactcaagcattccg 1991  
Qy 1986 CCATTGAGTGGCGGTTATGTCATAGAGAGAAACCTTACTTGTTCAGATTAAGATTTCT 2045  
Dd 1992 acataaaagctagcgctatattgttggaggaagctcgaagaatttctcgaaaaatctct 2051  
Qy 2046 TAGTCTGCAATTTTTCAGCTCTTGTGAAGAGATAGAACACTTGTGAGCTAACAATCA 2105  
Dd 2052 cagtgtagcttctcgcagctctctcgtgaagataaagacgtgttaatagtgtaaaataac 2111  
Qy 2106 AGGTACACTTACGAGAGAGACTCTTATACAGACAAACGAACCTTATATCTCTTCC 2165  
Dd 2112 ctctcataactatttagagctgctactacactgcaacatacgaagcattccctaaggagctcc 2171  
Qy 2166 TTGCAAACTAGCGGCTTGTGCTTGTCTTATGTTCTTACAGAGATTCCTGTTCTTTTC 2225  
Dd 2172 ---catgcccatttggaaatgataccgcgaatgctgaagaatctctctcatcttga 2228  
Qy 2226 AGGAACCTTACTACCTACCCATACGGAATACGATGTGAAGAACAAAGTATACAACTATCC 2285  
Dd 2229 tgcacagctaaagctacagctacatacaataatgatatgatactcgtataactctcacc 2288  
Qy 2286 TACTGTAAAGGAAGCTGGGGGAATGATAGTTTGCCTTAAAGATCGGTGGA---GAGC 2342  
Dd 2289 tgaagctcaaggtctcttgcacaaataatctcgggctctgaagctcgggaagatctctcgc 2348  
Qy 2343 TCCGATTTGCTTAGTGAAGTGTCTATTTGAGCAGCATACGCTTCATGAAATGCA 2402  
Dd 2349 tctatatctccctaagaagcaacgctctctccagggatattccctctcttaagttcca 2408  
Qy 2403 GTTTGTCTATGCACATCGAGGAAGTTTAAAGAACAGGGAACAGAGCTCGTAATTGG 2462  
Dd 2409 ggcagctctacagccgcacaacaaacttaagaagagtgcgctgaagccgctgctttga 2468  
Qy 2463 AAGTAGCGGTGTGATCTTGCCTTACCTATGCGGGAATCGCATGTAAGGAATGCA 2522  
Dd 2469 tgatggagacctgaaggaactgctctacatccctgcgcgactcgggttagaanaaatccga 2528  
Qy 2523 CTGCCAAGATGCAAGCTCAATCTAATCTTGGTTATCTGTGATCTCTTCTGATGTA 2582  
Dd 2529 agatgaaanaataatcttcgagattctctcagccaacatgttggtgtagtctcgtaaaaa 2588  
Qy 2583 CCCCAGCTGTACGACACAGCTGCGAATTAAGCGGTGATTTCTTGGAACCTTCGGTACGA 2642  
Dd 2589 tcccgcgtcgactctctctaatgctagtgagagccctctgagactcgtatgtaaaaa 2648

```

Db 1982 gtcttcgcaacaatgaatgaggtatggtgtagagcgacatacaattagcttcgata 2041
QY 2039 AGATTCTTGTCTGCTGATTTTGTGACGCTCTTTGGAGAATAGACTCTTTGTAGCTTA 2098
Db 2042 attctacactgcagctcttcgcaattatcggaaagaatagatactcattataata 2101
QY 2099 AGAATCAAGTACAGTCTCGGAGAACTCTGATTAACAGCAACGAACCACTATATCT 2158
Db 2102 aaatagagctctgccttcgacgtctctcattccagactcagaagctgctctt 2161
QY 2159 CTCTTCCTTGCAAACTACGCGCTTGCTGTGCTTATGTTCTTACAGAGATTCCTTC 2218
Db 2162 ctccaagctgttac-----gtaccttcctgtagatctgaagtgagcagcgctcc 2212
QY 2219 TCTTTTACAGAAACCTTACTACCCATACGGATACAGATCTGAAACCACTATACAA 2278
Db 2213 tctttagtctcagatcagctatctatctagtaaaatactagtaaaacctattacaccc 2272
QY 2279 CATATCTACTGTTAAGAGAGCTGGGGAATGATAGTTTCGCTTACAGAAATTCG---GTG 2335
Db 2273 aagcaccaagaagagagagctcgtgtataatgaagcgttcgctcgcgaactcgagct 2332
QY 2336 GAAGAGCTCCGATTTGCTTATGATGAAAGTCTCTATTTAGAGCAGTACATGCCCTTCATGA 2395
Db 2333 cctaccacacactgctttaaagcagatgagctctctccacggtattcttcctcatca 2392
QY 2396 AATTGAGTTTGTCTATGACATCAGCAAGGTTTAAAGAAC---AGGCAACGAAGCTC 2452
Db 2393 aagtagaagcttcgtacatacccaagaatagcttcaagaactaactaactctgttac 2452
QY 2453 GTGAAATTTGGAAGTAGCCCTCTTGTGAATCTTGCCCTTATCGGATCCGATTTGATTA 2512
Db 2453 gatcttcgtagatcggtgtatttaataacgtctctgctcattgaaattactctgaga 2512
QY 2513 AGGAATCAAGCTGCAAGATGCAAGCTACATCTAATCTTGGTTAATCTGATCTTG 2572
Db 2513 gattcccgagaacgagcgctgctctacgaagctactgctacatcagtcgagatgct 2572
QY 2573 TTGCTGTAAACCCGCTGTAGAGCAACCTGCAATTTGCGGTGATTTGTTGAAAACCT 2632
Db 2573 atcgttaagaatcccgagctcagcagcgtctctcctacacaaactcgttggaataca 2632
QY 2633 TCGGTACGAATTTGGCAGACAAAGCTTTAGTCTCTGCTGTCAGGGAACCAATTTGCTTAA 2692
Db 2633 caggaaacgaaatctctcaagaagcgtgatacggaagagcagagatcttattgctct 2692
QY 2693 ACTCAAAATTTTGAAGCCTTTAGCCAAATTTTCTTTGAATTTGCTGGGTCATCTGCAATT 2752
Db 2693 ctccaactcttgagtcacaagtaacctatctatggaatctcgtagatcttcacgagct 2752
QY 2753 ACAATGTAGCTTGAAGCAAAATACCAATTC 2784
Db 2753 acaatgagatcttgagtgaggtccagctc 2784

```

```

PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98WD-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
PI (CHRI/) CHRISTIANSEN G.
PI Birkelund S, Christiansen G, Knudsen K, Madsen A,
PI Myind P;
DR WPI: 1999-105610/09.
XX P-PSDB; AAW88424.
XX
PS Claim 6; Page 62-63; 115pp; English.
XX
CC This DNA sequence codes for the novel 97.6 kDa surface exposed
CC protein Omp11 (see AAW88424) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see AAW06816-27) encoding
CC Omp1-Omp15 proteins (see AAW88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
CC the other, and encode polypeptides of about 89.6-100.3 kDa and
CC about 56.1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp1-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.
CC
SQ Sequence 2793 BP; 803 A; 681 C; 567 G; 742 T; 0 other;

Query Match 18.4%; Score 514; DB 20; Length 2793;
Best Local Similarity 52.7%; Pred. No. 1,6e-139;
Matches 1404; Conservative 0; Mismatches 1200; Indels 60; Gaps 11;

QY 150 TGGAACTACTCTACCTTTAAGGGAATGTCAGAAATATTCGGAACAGGCAC 209
Db 162 tggaaagaaatgcttctatcagaagaatgctatata---aaagatgctgggaagagcac 218
QY 210 AGCAATCACAAAAAGCTGTTTAAACAACATAAGGCGATTTGACTTTTACAGATAACGG 269
Db 219 agcattaacagctgctgcttcaagaagaactaagggtgtagctgaacttacttgaaagg 278
QY 270 GAACCTCTATTTGTTCAAAAGGTGATGACGAGGACTGTAGCAAGGCGCTGCTTTAACAG 329
Db 279 atactatcttcaatcaacaagtagatgcggttctgaatgcagagcgcg---caag 335
QY 330 CAGCGTGNAGATTAATCTACACGCTTTAAGGCTTTTCTGCTCTCTTTTTCGCTG 389
Db 336 cacaactgctgaataaagcctcaactacacagatattctcaactcttcctcattgcagc 395
QY 390 TCTGTGAAGTTGATTAACACGCGCAAGAGAGCGCTTAGCTCTACAGGATGAGTTGAAG 449
Db 396 tcttggaactaacagtgcttcaggaataagtaacttaa---gttcgagggagccttaaa 452
QY 450 GTTTGACAAAATGTCAGTTGCTTTCAGCAAAAACCTTTCAACGG-----ATTA 500

```



```

104297 CTGATGGGATCACTATCAATTAATCTTGTCTCAATGATGATTCCTTAAAGAGACCAAG 104238
1511 AGGCAAAATAGAAACCAAGCAGCTACCAAAAATCTGATTTATGAGACATCACTT 1570
104237 AGGCTAGCGCTGAAAAGAACCAAGCAAGTCAGACAGCTTATATCTGATGCTCTGTC 104178
1571 TATGAGACCCGAGGCGAGCTTTATGAAATCATAGTTTAAAGAAATCCCTACGCTCAG 1630
104177 TTGTATATCTCTTGGAATATGTCATGAGATATCTCTTGGAATACCCCTCAAGTCTTT 104118
1631 ACATCTTAAAGCTCAAGCTT-----CTGGAACCTGTAAACAGACCGCAGTACTCCAG 1684
104117 CTGTCTCACTCTTACTGCTGAGACCCCGCAGATATTCACATCAAGACTTGTAGCTGTC 104058
1685 ATCCCTAATATGGGTGAGAAATTCATTAAGGCTATGAGGAATTTGGGCCCAATGTTT 1744
104057 ATCCCTAGAAAAAAATCTATCCATTTGGGATGCCAAGGAAATGCGCATTAATCTTGGC 103998
1745 GGGGACAGGGGC-----TTTACGACTGCAACCTTCAACTGACTAAACGTGCTATA 1798
103997 AAGAGATATCTGCGACTAATATCCAAAGCAGCGACTTCTACTGCAAAAACAGATACA 103938
1799 TTCTTAATCCCGAGCGTATGCGCTTTTATGCTTATAGCTTATAGCTTATATAG 1858
103937 ATCCGATCTCGAGCGCTGCGTGAACCTTAGTTGCTAACACGCTATGGGATCCTTTGTTG 103878
1859 ATATTTGCTCTCCCAATTTATCTTATGAGACTGCAACGAAAGGTTGCGAGGAGACCTG 1918
103877 ATGTGGGCTGCACACACAGCTGTAGCCACTAAATGACCCCAATCTCAAGAAACCTGGC 103818
1919 CTTTTGGTGTGCTGATTATCTTAATCTCTTCCATAGATAGTATACAAAACAGCAGCG 1978
103817 GCATCTGTGTGAGAGATCTGAACTCTTCCATTAAGTATACAGAAATGAATAAAG 103758
1979 GGTTCGCCATTTGATGCGCGTTATATGTCATAGAGAAACCTACTTGTTCAGATA 2038
103757 GTTTTGCGCACATTAAGTGAAGTATGTGTAGAGACGACTACACATTAAGCTTCGATA 103698
2039 AGATCTTATGCTGATTTTGTCTGAGCTCTTTGAGAGATAGAGACTCTTTAGCTA 2098
103697 ATCTTATCACTGAGCGCTTCTGCAATTTATCGGGAAGTATAGATCACTTTATTAATA 103638
2099 AGAATCAAGGTACAGTCTACGAGAGACTCTTATTAACAGACCAAGAAACCTATATCT 2158
103637 AAATAGAGCTTGTGCTATGAGAGCTTCTCTCATCTCCAGACATTAAGGAGACTTCTCT 103578
2159 CTCTTCTTGCAGAACTAGAGGCTTGTCTGTCTTATGTTCTTACAGAGATTCCTGTC 2218
103577 CTCCAAGCTGTGTAC-----GCTACCTTCTGAGTCTGAAGTGAAGCAGCCTGTCC 103527
2219 TCTTTCAAGAAACCTTACTACACCCATAGGATAGATACATCTGAAACCAAGTATACAA 2278
103526 TCTTTATCTGATGATCACTATCTATATGTAATAAACTATGAAACCTATTTACACCC 103467
2279 CATATCTACTGTTTAAAGAGAGCTGGGGAATGATGTTTCTGCTTATGAAATTCG---GTG 2335
103466 AAGCACCAGAGGAGAGAGCTGCTGTATATATAGCGTTGCGCTCGAAGCTTGCAGCT 103407
2336 GAAAGACTCGATTTGCTTATGATGAAAGTCTCTATTTGAGACATGACCCCTCATGA 2395
103406 CCTTACACACACTGTTTAAAGCATGAGGAGTCTTCCACGCGTATTTTCTTTCATCA 103347
2396 AATGAGATTTGCTATGACATCAAGAGGTTTAAAGAC---AGGAAACAGAGCTC 2452
103346 AAGTAAAGCTTGTGATACATACCAAGATAGCTTCAAGAGATTAATTAATCTCTGTGATC 103287
2453 GTGAATTTGGAAGTACCGCTTGTGATCTTGTGCTTACTTATCGGAGATCCGATTTGATA 2512
103286 GATCTTTCGATGAGGCTGATTTATTAAGCTGCTGCTATTTGGAATTTACTTGGAGA 103227
2513 AGAATCAGACTGCCAAGTGCAGCTACATCTTAACCTTGTGTTATCTGATGATCTTG 2572

```

```

103226 GATTTCTGAGAAAGAGCGTCTTACGAAAGCTACTGTCATCTACGTTCCGATGCT 103167
2573 TTGTAATACCCCGACTGATAGCAACACTGGCAATTAAGCGGTGATTTGGAAAACT 2632
103166 ATGTATAGATCTCTGATGACACAGCTTCTTATCAACATTAAGCTGGAAGAACTA 103107
2633 TCGTACGAATTTGGGAGACAGCTTTATGCTCTGTGCGAGGAAACATTTTGTGCTTA 2692
103106 CAGGAAGCATCTCTCAGACAGCTGTATCGAAGAGAGGAGATCTTTATGCCCTCT 103047
2693 ACTCAATTTTGAAGCTTTAGCCATTTTCTTTTAATTTGCTGCTGCTATCGCAAT 2752
103046 CTCCAATCTTGAAGTCTACAGATTAACCTATCTATGAAATTCGTGATCTTACGAGCT 102987
2753 ACAATGATAGAGAGCAAAATACCAATTCAT 2787
102986 ACATGACAGATCTTGGAGGATTAAGTTCCAGTTCTAA 102952

```

## RESULT 12

AAA30852  
ID AAA30852 standard; DNA; 2784 Bp.

AC AAA30852;

DT 29-AUG-2000 (first entry)

XX Chlamydia antigen CPN100638 gene open reading frame.

XX Chlamydia antigen; diagnosis: infection; community acquired pneumonia;

KW therapy; upper respiratory tract disease; bronchitis; sinusitis;

KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;

KW ds.

XX Chlamydia pneumoniae.

XX Key Location/Qualifiers

EH CDS 1..2784

FT /tag= a

FT /partial

FT /product= Chlamydia antigen CPN100638

FT /note= "no stop codon given"

XX WO200032794-A2.

PD 08-JUN-2000.

XX 01-DEC-1999; 99WO-CA01147.

XX 01-DEC-1998; 98US-0110339.

PR 01-DEC-1998; 98US-0110340.

PR 01-DEC-1998; 98US-0110427.

PR 01-DEC-1998; 98US-0110428.

PR 01-DEC-1998; 98US-0110438.

XX (CONN-) CONNAUGHT LAB LTD.

XX Murdin AD, Oomen RP, Wang J;

XX WPI; 2000-412339/35.

DR P-PSDB; AAY90239.

XX Nucleic acids encoding polypeptide antigens from chlamydia useful for

PT preventing, diagnosing and treating diseases such as community acquired

PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset

PT asthma.

XX Claim 2; Fig 5; 174pp; English.

CC This sequence encodes a Chlamydia antigen of the invention, designated  
CPN100638. The nucleic acids (and their complementary sequences) may be  
used as diagnostic agents for detecting the presence of nucleic acids  
encoding Chlamydia antigens in samples according to standard methods,

AC AAC81914;  
XX 27-FEB-2001 (first entry)  
XX Chlamydia pneumoniae genome DNA.  
XX Genome; diagnosis; vaccine; ds.  
XX Chlamydia pneumoniae.  
XX W0200027994-A2.  
XX 18-MAY-2000.  
XX 12-NOV-1999; 99MO-US26923.  
XX 12-NOV-1998; 98US-0108279.  
XX 08-APR-1999; 99US-0128606.  
XX (REGC ) UNIV CALIFORNIA.  
PI Stephens R, Mitchell W, Kalman S, Davis R;  
XX WPI; 2000-376516/32.  
XX  
XX Isolated nucleic acid for use in diagnostic and analytical methods  
PT encodes genomic sequence of Chlamydia pneumoniae -  
XX  
XX Claim 2; Page 128-320; 320pp; English.  
XX  
XX This invention describes a novel nucleic acid (NI) encoding a Chlamydia  
CC pneumoniae protein (PI), given in the specification. The isolated nucleic  
CC acid is useful for diagnostic and analytical methods, such as,  
CC hybridization-based assays or amplification-based assays. The protein may  
CC be used for diagnostic purposes, for their enzymatic or structural  
CC activity, or as a vaccine. The invention also describes (1) a probe  
CC comprising a hybridizing fragment of NI; (2) an isolated nucleic acid  
CC (N2) that hybridizes under stringent conditions to NI; (3) an expression  
CC cassette comprising NI under the transcriptional regulation of a  
CC transcriptional initiation region functional in an expression host, and a  
CC transcriptional termination region; (4) a cell comprising an expression  
CC cassette of (3) as part of an extrachromosomal element or integrated into  
CC the genome of a host cell as a result of induction of the expression  
CC cassette into the host cell, and the cellular progeny of the host cell;  
CC (5) a method for producing a PI comprising growing a cell of (4) where  
CC the protein is expressed and isolating the protein free of other  
CC proteins; (6) a purified polypeptide composition comprising at least 50  
CC weight % of PI; and (7) a monoclonal antibody binding specifically to the  
CC peptide of (6).  
XX  
XX Sequence 273254 BP: 76423 A; 51054 C; 61965 G; 83812 T; 0 other;  
SQ

Query Match 18.68; Score 517; DB 21; Length 273254;  
Best Local Similarity 53.18; Pred. No. 2.3e-139;  
Matches 1420; Conservative 0; Mismatches 1180; Indels 75; Gaps 12;

QY 143 AAGATGCTGGAACCTACCTATTTAAGGAATGTCACCTAGAAATTTCTCTGAA 202  
DB 105581 ATGAGAGATGGAACCTATATATATCAACGAGGATGTC-- -AATCCCAATGACAT 105525  
QY 203 CAGCAGCAGCATCAAAAAGCTGTTTAAACACACTAAGGCGATTTGACTTTACAG 262  
DB 105524 CTGCGACAGCTCTACCGCTTCGCTTTAAAGAACTACGGAATTTCTTTCCAA 105465  
QY 263 GTAACGGGAAGCTCTATTTGTCAAAACGATGATGACGAGGAGCTGACAGGGGCTGCTG 322  
DB 105464 GCCACGGCTACCAATTTCTCTCAAAAATATGATGCGGG-- -AGCGAATCTACT 105411  
QY 323 TTACACAGAGCTGTATTAATCTACACGTTTATAGGTTTCTTCGTACTTCTTTA 382  
DB 105410 TTACCAATACAGCTGCAAAATAGCTTCTCTTTCAGGATTTCTCTATTTGTCACATA 105351

QY 383 TTGCGTCTCTGGAAGTTGATTAACCTACCGCAAGAGCCGTTAGTCTCTACGGTA 442  
DB 105350 TAC-----AACCACGAATGCTACACAGAAACAGAGCCATCAAGTCCACAGAG 105300  
QY 443 GCTTAAGTTTGACAAAATGTCATTTGCTCTCAGCAAAAACCTTTCAACGATTAATG 502  
DB 105299 CTGTCTATTCAGTCGACATATAGTGTCTACTTTGGCCAAAACCTTTCTATGCAATG 105240  
QY 503 GCGGTGCTATACCGCAAAAACCTTTCTATTAACGAGGACTACATGTCAGCTCTGTTT 562  
DB 105239 GAGCGCCCTCCAAAGCGAGCTCTATCATCTATGCTTAACCCCACTAAC-- -GTTTG 105183  
QY 563 CTGAATATACCTCTCAAAAGAAAGCGAGCCATTACAGCTTCGATCCCTTACATTA 622  
DB 105182 CCAAAAACAAAGCAACGCAAAAAGGGGGTGCCTCTATTCCACGGAGGAGATTACATTA 105123  
QY 623 CTGGAACCAAGGGGAATCTTTTCTGACATATCTTTCCGATTTCTGAGCTGCA 682  
DB 105122 ACAATACGTTAACTCAACATCATTTTCTGAAAATACCGCGCAATGCGGAGCA 105063  
QY 683 TTTTACAGAGCCCTCGGTGACTATTTCTAATATGCTAAAGTTCCCTTATGACATA 742  
DB 105062 TTTACACGGAAGCTAGCATTTTATTAGCACCAACAAATTAAGCTTTAATAACAATA 105003  
QY 743 AGGTACAGAGCGAGCTCTCAACACGGGAGATGTCAGAGGTCTATCTGCTT 802  
DB 105002 GTGTACCG-----CAACCTCAGCTACAGGGGAGCCATTACTGTACTA 104958  
QY 803 ATAAATAGTACAGATCTAAGGTACCTCTCAGTGAATTCAGATGTTACTCTTACGA 862  
DB 104957 GTACATCAGCCCCCAACCACTCTTAATCTATCAGAACAGGGCACTGAACCTTATAG 104898  
QY 863 ACAATACATCAGCAACACAGCGGAGAGTATCTATGTAAGAGTGAACCTGCTCGC 922  
DB 104897 GAATACAGCAATTTCTAGTGAGGAGGATTTACTGACAAATCTATTTCTTTCTG 104838  
QY 923 GAGCACTTACCTATTCAGTAAGTAAGTGTCAATGAGATGACGCTCTAAGAGTGAG 982  
DB 104837 GAGACCTACGCTTTTAAACAACTGCTGATATGATACGACGCTCTCTAGGAGAG 104778  
QY 983 CCAATGCTATGAAAGATGAGGAGATTTGATTTATCCCGGATATGCTGATGCT 1042  
DB 104777 CAATGCTATGCTGACCTCTGATCTTGAATCTTTCGCTTGTGGAACATCTACTT 104718  
QY 1043 TTTTGAATTAACGT-----CACTTACTACTCTCGGGGCAATTAAGTAGTA 1093  
DB 104717 TTGAAGGAACACAGTATGTAAGAGAGCTTTCTGAGTCAACACCTACGAAATTTTA 104658  
QY 1094 TCGACTTAGA-- -ACGAGTCAAAAGATGACAGCTTTGCTGTGCTGCTAGAGCA 1150  
DB 104657 TTAACATCGGAACACCAATGCTAAGATTGTACAGCTCGAGCCCTCAAGCAATTA 104598  
QY 1151 TCTACTTCTATGATCCCTAATCTACAGATTTCCACACAGCTTACAGATGTCTTAAG 1210  
DB 104597 TCTACTTCTATGATCCCTAATCACTACATCACTGACGCTCTCTCAATCTCTTAAC 104538  
QY 1211 TTAATGACATCCGCGACATTTCTGACATTAATATACAGGGAATCTCTACAGAG 1270  
DB 104537 TAAATGCTCTGACCTTTCAGAGGAATCTGCAATTCAGAGAACCATGCTATTTCTGAG 104478  
QY 1271 AAAAGTATACAGACAGAGCCGAGATTTCTAATAATCTTACTCTGAAAGCTACTACAGC 1330  
DB 104477 AGAAGCTCTCGGAAGCAAGACCTGAGAAAGTGATTAATCTCAATATTCATTAAGCAAC 104418  
QY 1331 CTGTACTCTTTACAGAGAGTCTATCTTTAATAACATGAGTACTCTGACACTCAGG 1390  
DB 104417 CTGTACTCTTTGCGAGGAGGAGCACTCTCTTAATCAAGGAGTCACTTACTTCTAAGT 104358  
QY 1391 CATTCACACAGAGCAATCTGCTGCAAAATGAGAGTGTGGAATCTCTAGACCTG 1450  
DB 104357 CTTTTCGCAATTCCTCGGCTCTACCTCTCATGATGACAGGAGCAATTAAGAACCG 104298  
QY 1451 CTGATACTAGCACCATTAACAATTTGTATTAACATCAGTTCTATAGACGCTCAAAA 1510

QY 743 AGGTACAGAGAGGAGCTCTCTCAACAGGGGATATGTACAGAGGTGCTATCTGTCTT 802  
Db 737 ggtgagccg-----caactcagctacaggggagccatttactgttagta 781  
QY 803 ATTAACCTAGTACAGTACTAGGTCACTCCCTCACTGGAAATCAGATGTACTCTTCAGCA 862  
Db 782 gtacacagcccccaaacagcttctacatctacacagcgggaacccaatttag 841  
QY 863 ACATATCATCGACAAACAGGGGAGAGCTATCTATGTGAAAAAGCTCGAACTGGCTCCG 922  
Db 842 gaatacagcaatctactagtggtggcgatctactacagcaatcagttcttctctg 901  
QY 923 GAGAGTATCCCTATTCACTAGAAATAGTCTCAATGAGAGTACAGTCTCTAAAGTGGAG 982  
Db 902 gaggactacgcttcttaaaaaaactcgtctataagactcagctcccttagggagg 961  
QY 983 CCATACCTATCGAAGTACTGGGGAATTGAGTTATCCGCCATYAGTGTGACATTGTCT 1042  
Db 962 caatgcgattcgtcagctcgtgactcttgccttccgctctcgtggtgagacatact 1021  
QY 1043 TTTTATGGAATACAGT-----CAGTTCTACTCTCTCGAGAGATAGAAAGTATGTA 1093  
Db 1022 ctgaaggaacacagtagtcaaaaggagcttctcgaagtcagaccactacacgaatctca 1081  
QY 1094 TCGACTTATGAG--ACGAGTGCAGAAAGATGACAGCTTTGGCTTCTGCTGTGTAGAGCCA 1150  
Db 1082 ttaacatcggaaacacaaatcgtcaagattgtacagctcgtgagctctcgaagaaatacta 1141  
QY 1151 TCTACTTCTATGATCCCATTACTACAGGATCTTTCACAAACAGTTACAGATGTCTTAAAG 1210  
Db 1142 tctactctatgctcctacacaaacacatctcagctcctcctcagctcctcaact 1201  
QY 1211 TTTATGAGACTCCGGGAGATTTCGACTCAATATACAGGAAATCATCTTACAGGAG 1270  
Db 1202 taaatgctcgcagcttgcaggaatcctgcacatacaagaacacatcgtattctcggag 1261  
QY 1271 AAAAGTATAGAGACAGAGGCGCAGATTCTTAAATCTTACTTGAAGCTACTACAG 1330  
Db 1262 agaatcctcggagagagaagctcgaagaagctgataatctcaatctacatcagcaac 1321  
QY 1331 CTGTAACTCTTTCAGAGAGTACTCTTCTTTTAAACATGAGTACTGTGACAGTCAAG 1390  
Db 1322 ctctaactcttgcgggagcaactctccttaacacagagatcactagctagtgtaagt 1381  
QY 1391 CATTCACTGAACGAGGAGATTTCGCTCGAATAATGACGTAGCAATCTCTGAACCTG 1450  
Db 1382 ccttctcgaatctccgggctcctacccctcctatgtagcaggagacacattagaacaag 1441  
QY 1451 CTGATCTAGCACCATTAACAAATTTGGTATTAACTATGAGTTCTATAGAGGTGCAAGA 1510  
Db 1442 ctgtagtgatcactatcaataatctgtctcctcaatgtagatctccttaaaagagacaaga 1501  
QY 1511 AGGCAAAAATAGAAACCAACAGCTACGTCAAAAATCTGACTTTATGTGAAACATCACT 1570  
Db 1502 agagctacgcttaaaagaacaaacagagcagcagcactctatctatctgtagctctc 1561  
QY 1571 TATTGACCCGAGCGGACGCTTTTATGAAATCATGTTTAAAGAAATCCGATCTCTAG 1630  
Db 1562 tctgtagatcctctcgtgaatgctacgaagatgctcctcgtgaataacctcaagctctt 1621  
QY 1631 ACATCTTAGAGCTCAAAAGCTT-----CTGGAACCTGTAACAGACCGCAGTACTCAG 1684  
Db 1622 ctgtctcactcttactgtcagcagcccgagataatcaatcaacagacttagcgctg 1681  
QY 1685 ATCCCTAATATGGTGAATTCATTAAGGCTATAGAGGAACTTGGGCCCAATTGTTT 1744  
Db 1682 atcccttaaaaaaaatctcatctcattcgttggaataccaaggaatctggaatctctg 1741  
QY 1745 GGGGGAAGAGGCG-----TTCTACGAGTCAACCTTCAACTGAGTAACAGTGGCTATA 1798  
Db 1742 aagagagatctcgtactaataatccaagcagcgactcttctcctcgtgacaaaaaagataca 1801

QY 1799 TTCTAATCCGAGCGATATCGGCTCTTTAGTCCCTAATAGCTTATGAAATGATTAAG 1858  
Db 1802 atccgaatctcgtcgctgtagaacttagtgctaaacagcttagggatcccttgctg 1861  
QY 1859 ATATTAGCTCTCTCCATTATCTTATGAGACATGCAACGAAGGTTGCGAGGAGACCGTG 1918  
Db 1862 atgtgcctccacatacaacagcttgtagccactaaagtagcccaatctcaagaacctcg 1921  
QY 1919 CTTTGTGTGCTGATTAATCACTTCTTCATTAAGGATTAATACAAAACGAGCGG 1978  
Db 1922 gcaatcgtgtggaagagatccgcaactctccataaagaatagacgaataataaag 1981  
QY 1979 GGTTCGCCATTATGAGTGGGCGGTTATGTCTATAGAGAAACCTACATCTTGTTCAGATA 2038  
Db 1982 gtttcgcaacataagtgcaagtgtagtgtaggagcgactaacaacttagctcgata 2041  
QY 2039 AGATTCTTAGTGTGCTGATTTTGTACAGCTCTTTGGAAGAGATAGAGACTTCTTGTAGCTA 2098  
Db 2042 atcttactcagcagcctctcgtcgaattatcgggaagataagatacacttataata 2101  
QY 2099 AGAATCAGGTACAGCTTCAGGAGAACTCTATTAACGAGCAACGAACCTATATCT 2158  
Db 2102 aaatagagcttctcgtctatgcagcttctcctcaatcccaagatcagcagcttct 2161  
QY 2159 CTCTTCCTTGCAAACTACGCGCTTGTCTGTGCTTATGTCTTACAGAGATTCTGTTC 2218  
Db 2162 ctccaagcttgtagc-----gtactcctctcgtatctgtaagtgagcagccgtgc 2212  
QY 2219 TCTTTTCAGGAAACCTTACTACCCCATACGATATACGATCTGAAACCAAGATATACAA 2278  
Db 2213 tctttagtctcagacacatataatctataagtaaaatctagaaacactatacacc 2272  
QY 2279 CATATCTCTCTGTTAAAGAGAGCTGGGAGATGATGTGTTCCGCTTAAAGTTG--G 2335  
Db 2273 aagcacaagaagagagagctcgtgtgataaiaagcgttgcgtctcggaaatcgtgagct 2332  
QY 2336 GAAGAGCTCCGATTTCTTAGATGAAGAGTCTCTATTGAGCATGACATGCCCTTACGA 2395  
Db 2333 ccttaccacaacacttgcttaagcagtagggtctcttccaagcttagtttcccttaca 2392  
QY 2396 AATTGCAATTGTCTATGCAATCAGAGGAAGTTTAAAGAAC--AGGAAACGAAGCTC 2452  
Db 2393 aagtagaaaccttctgatacaacaagaatagcttcaagaagcgtatatacacttctg 2452  
QY 2453 GTGAATTTGGAAGTACCGCTGTGTAATCTGCTTACTATCGGATCCGATTTGATA 2512  
Db 2453 gatcttctgatacggtgatatataaagctcgtcgtcctatgtaataatcaactcgaag 2512  
QY 2513 AGGAATCAGACTCCCAAGATGCAACGTACAAATCTATCTTGGTTTACTGTGATCTTG 2572  
Db 2513 gatctcggagaagcggtcgctcttaagaagctactgtcaactcgttgcgagtgct 2572  
QY 2573 TTGCTAGTAAACCCGACTGTAGACAACACTGCGAATTAGCGGTGATTCTTGAAAACT 2632  
Db 2573 atcgttaagaatcctcgtcagcagcagcttccctcaatacaatlaactcgttggaaacta 2632  
QY 2633 TCGGTACGAATTTGCAACAACAAGCTTTATGCTTGTGTCAGGAGCACTTTTGGCTTA 2692  
Db 2633 caggaaagcatctcttaagaacaaagctgtagatcggaagagcaggtactttagcttct 2692  
QY 2693 ACTCAAAATTTGAAGCCTTTAGCCAAATTTCTTTGAATTTGCGTGGTCTGCAATT 2752  
Db 2693 ctccaactcttgagtcacaagaatcaactctatagaaatctcgtatcttcaagagct 2752  
QY 2753 ACAATGTAGCTTAGAGCAAAATACCAATTCTAA 2787  
Db 2753 acaatgcagatcttggaggtgaagttccagttctaa 2787

RESULT 11  
AAC81914/C  
ID AAC81914 standard; DNA: 273254 BP.  
XX



```
Db 2314 cctcgatttagaagggcagctcgcttatagccagctcaatgatgctgaagacaaga 2373
Qy 2274 TACACATATCTCTACTGTTAAAGAGCTGGGGAGATAGTTTGGCTTAAAGATTGG 2333
Db 2374 tactgcgatactcgtggaagagcttcttgagggaataatgctttaacagatgttgg 2433
Qy 2334 TGAAGAGCTCCGAGTTTCTTAAAGATGAAGTGCCTATTTAGAGCATCATGCTTCAT 2393
Db 2434 agctctctcattctctctcgaatacctgcatgttcttgaacctgtctcctacat 2493
Qy 2394 GAAATTCAGATTGTCTATGACATCAGGAAGTTTAAAGAACGGAGACAGAGCTCG 2453
Db 2494 caaatgactcgaacctatacgtcagagacgtcttcgagaaaggtacagaagagag 2553
Qy 2454 TGAATTTGGAAGTGCCTCTGTGAATCTTGCTTACCTATCGGAGATCCGATTGATTA 2513
Db 2554 atcttcttgacagcaacctctcaatcttctgctctataggggtgaagttgagaa 2613
Qy 2514 GGAATCAGACCTGCAGAGATGACATGATCTTGTATGATGATCTTGT 2573
Db 2614 gtctctgattgtaagtcttctctatgactgacttaccatgtctcctgactctat 2673
Qy 2574 TCGATGATACCCCGACTGACAGACACTGCGAATTAAGCGGTGATTCTTGAAAACTT 2633
Db 2674 ccgaatgaccccaatgcactacagcacttgtaatcagcgagcctcttggaactta 2733
Qy 2634 CGGATACGAATTTGGCAACACAAAGCTTGTGCTTGTGCTGAGGAGCAATTTTGTCTTA 2693
Db 2734 tgcgaataactagcagcagcagcgcttgcaagtgctgagcgactcactcgcctctc 2793
Qy 2694 CTCAAAATTTGAGCGCTTAAAGCTTAAATTTGAAATGCGTGGATCTCGCAATTA 2753
Db 2794 tccatgttctgaaagtgcgcgcagcttgccttgaagttcgtgactcctcagattta 2853
Qy 2754 CAATGTAGCTTAGAGCAAAAATACCAATTTCTA 2786
Db 2854 taatgtatcttg99gtaagtccaatctta 2886

RESULT 10
AAK06822
ID AAK06822 standard; DNA; 2787 BP.
XX
AC AAK06822:
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp10 DNA.
XX
KW Omp10; outer membrane protein 10; surface exposed protein; antigen;
KM infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX
OS Chlamydia pneumoniae.
XX
PN WO9858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98WO-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.
XX
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX Myind P;
XX
DR WPI: 1999-105610/09.
DR P-PSDB: AAW88423.
XX
PT Species-specific test for identifying mammals infected with
Chlamydia pneumoniae - comprises detecting antibodies specific for
```

```
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT these proteins
XX
PS Claim 6; Page 59; 115pp; English.
XX
CC This DNA sequence codes for the novel 98.4 kDa surface exposed
CC Chlamydia Omp10 (see AAW88423) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see AAK06816-27) encoding
CC Omp4-Omp15 proteins (see AAW88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12, 11, 10, 5, 4, 13 and 14 in one cluster and Omp6, 7, 8, 9 and 15 in
CC the other, and encode polypeptides of about 89, 6-100, 3 kDa and
CC about 56.1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.
XX
SQ Sequence 2787 BP; 815 A; 689 C; 535 G; 748 T; 0 other;

Query Match 18.6%; Score 517; DB 20; Length 2787;
Best Local Similarity 53.1%; Pred. No. 2,1e-140;
Matches 1420; Conservative 0; Mismatches 1180; Indels 75; Gaps 12;

Qy 143 AAGATGCTGGAACCTACCTATTTAAAGGAATGTCACTTGAATAATTTCTGCA 202
Db 158 atgcagatggaactatactataatcaacagggatgtctc--atacaccatgagagt 214
Qy 203 CAGGACGACATGCACAAAAGCTGTTTAAACACTTAAGGGGATTTGACTTTCACAG 262
Db 215 ctccgacagcctcaacgcgttcctcgtcttaagaacacactggaactcttctccaag 274
Qy 263 GTAACGGGACCTCTCTATTTGTTCCAAACGGTGTGATGACAGACTGTACAGGGCTGCTG 322
Db 275 gcaacggttacaactttctcctacaacaaatcagatgcgsg-----agcgactgtact 328
Qy 323 TTAAACAGCAGCGTGTGATATAATCTACCAAGCTTTATAGGCTTTCTTCGCTATCTTTTA 382
Db 329 ttaccaatacagctgcaataaagcttctccttccagatcttccctattgttcaactaa 388
Qy 383 TTGGCTCTCCGGAAGTTGCAATTAACCTACGGCAAGAGCGGTTTGTGCTACGGGTA 442
Db 389 tac-----aaaccaagatgctaccacaagacaagagacccaagtcacagagag 439
Qy 443 GCTTGAAGTTTGACAAAATATGCTAGTTTGTCTTCAGCAAAAATTTTCAACGATTAATG 502
Db 440 ctgtctcattcagtcgacatacagtgcttactcttggtgcaaaacttctctaatagcaatg 499
Qy 503 GCGGTGTATACCGCAAAAACCTTTTCATTAAACAGGAGTACATAATGTACACTCTGTTT 562
Db 500 gaggcgccctccaagcgactcactacgtatcgtacaaacccaactaac--gtttg 556
Qy 553 CTGAAAATATCTCTCTCAAAAGAAAGCGGAGCATTCACATTCGATGCCCTTACCATTA 622
Db 557 ccaaaaacaagaacaaagaaaggggtgccccttaccacgggaggaataacaa 616
Qy 623 CTGGAACCAAGGGAGGAGCTCTTTTCTGACAAATATCTTTCGATTCGAGAGCTGCA 682
Db 617 acaatacgttaactcagcatcatttctgaaataacccgcygcaacaatgcyggaagca 676
Qy 683 TTTTTCAGAAAGCTCGGTGACTATTTCTTAATATGCTTAAAGTTTCTTTATGCAATA 742
Db 677 tttaacggaagcagctgatttattagcagcaacaagaacattgacttataacata 736
```





```

Db 1842 aagaggtactgagcactaaatccaaagcagactcttactctgacaaacagatata 1901
QY 1799 TTCCTAATCCCGAGCGATCGGCTTTAGTCCCTAATAGCTATGAGANTGCATTTATAG 1858
Db 1902 atccgaatccctgagcgtctggaacctagctgtgtaacacgctatggggatccttggtg 1961
QY 1859 ATATTACTCTCTCGCATTTATCTTATGAGACTGCAACGAAAGGTTGCAGGAAACCGTG 1918
Db 1962 atgtagcgtccatacaaacagctgtgagccataaagtaagcgaactccaaacacgcgcg 2021
QY 1919 CTTTTCGTGCTGCTGATTAATCTTACTTCTTCATTAAGATAGTACAAAAACAGCAGCGC 1978
Db 2022 gcatctggtgtggaaggatcccgactcttcctcaataagataagcagagaagtaataag 2081
QY 1979 GGTTCGCCATTGATGAGTGGCGGTTATGTCATAGAGAGAACTCATATCTTGTTCAGATA 2038
Db 2082 gtttcgccacataagtgaggttatgtgtgagagcgacataacatlaagcttctgata 2141
QY 2039 AGATTCTTAGTGCTGATTTTGTAGCTCTTGGAGAGAGATAGACTACTTTAGCTA 2098
Db 2142 atctatcacctgagcgtcttcgcacatcttcggaagaagataagatcaacttataata 2201
QY 2099 AGAATCAAGGTACAGTCTACGAGAACTCTCTATACACAGACAGAAACTATATCT 2158
Db 2202 aaatagagcttcgtcctatgacgctctcctccatccacgacatctagcgacctgtcct 2261
QY 2159 CTCTCTCTTGGCAATCTAGCGGCTTGTCTGTATATGTTCTTACAGAGATCTCTGTT 2218
Db 2262 ctccaagctgtttac-----gtacactctctgagatcgaagtgagagcctgtcc 2312
QY 2219 TCTTTTCAGGAAACCTTAGCTACACCCATACGAGTAAGATAGTAACCAAGATACAA 2278
Db 2313 tctttgactgcacatcagcatatactacatagtaaaatacactagaaacattacacccc 2372
QY 2279 CATATCTTACTGTTTAAAGAGAGCTGGGGAATAGTATGTTTGCCTTTAGATTGCG---GTG 2335
Db 2373 aagcaccaaaggagagagctcgtgtgataatagcgtgtgcctcgtgaacttcgagcgt 2432
QY 2336 GAAGAGCTCCGATTTGCTTATGATGAAGTGTCTATTTGAGCAGTACATGCGCTTCATGA 2395
Db 2433 ccctaccacacacgtcttaagcatalgaggtctctctccacgctattctctcatca 2492
QY 2396 AATTGCAGTTTGTCTATGACATCAGAGAGGTTTAAAGAAC--AGGGAACAGACAGCTC 2452
Db 2493 aagtagaagcttcgtacatacaaccaagatagcttcaagaacgtaactaactcctgttac 2552
QY 2453 GTGAATTTGGAGTAGCCGTCTTGTGATCTTGCCTTACCTATCGGAGTCCGATTTGATA 2512
Db 2553 gactcttcgataagcgtgatttaataagctctcgtgacctatggaattacactcgaga 2612
QY 2513 AGAATAGACTGCGCAAGATGCAACGTAACATCTCTGTTGATATCTGTGATCTGTCG 2572
Db 2613 gatctcagagaaagagcgtgctgtcttaagactactgcatcttaagctgtgcagatcct 2672
QY 2573 TTGCTACTAACCCCGATGTAAGCAGACACTGCGAATTAGGGGATTTCTGAAAAACT 2632
Db 2673 atcgtagaactccgactgacgacgacgctccccaatacaaaacactcgtggaaaaacta 2732
QY 2633 TCGGTAGCAATTTGGCAAGCAAGCTTTAGTCTCTTCGTGAGGAGAACCACTTTGCTTTA 2692
Db 2733 caggaacgaactcctcaagaacgctgtatcgaagaagcagatcttcttaicctct 2792
QY 2693 ACACCAATTTTGAAGCTTATAGCCCAATTTCTTTGATTTGCGTGGGTGATCTCGCAATT 2752
Db 2793 ctccaactctgaggtacacgaactactatgaaactctgagctcttcacgcagcct 2852
QY 2753 ACAATGATAGCTTAGAGCAAAATACCAATTTTAA 2787
Db 2853 acaatgcaagatcttgagagtaagttccagttctaa 2887

```

RESULT 9

```

AAA30849
ID AAA30849 standard; DNA: 2950 BP.
AC AAA30849;
DT 29-AUG-2000 (first entry)
XX
DE Chlamydia antigen CPN100635 full length coding sequence.
KW Chlamydia antigen; diagnosis; infection; community acquired pneumonia;
KW therapy; upper respiratory tract disease; bronchitis; sinusitis;
KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;
ds.
OS Chlamydia pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 101..2887
FT /tag= a
FT /product= Chlamydia antigen CPN100635
FT sig_peptide 101..229
FT /tag= b
FT mat_peptide 230..2884
FT /tag= c
FT /note= "mature CPN100635"
XX
PN MO200032794-A2.
XX
PD 08-JUN-2000.
XX
PF 01-DEC-1999; 99MO-CA01147.
XX
PR 01-DEC-1998; 98US-0110339.
PR 01-DEC-1998; 98US-0110340.
PR 01-DEC-1998; 98US-0110427.
PR 01-DEC-1998; 98US-0110428.
PR 01-DEC-1998; 98US-0110438.
XX
PA (CONN-) CONNABUHT LAB LTD.
XX
PI Murdin AD, Oomen RP, Wang J;
XX
DR WPI: 2000-41239/35.
DR P-PSDB: AAY90237, AAY90238.
XX
PT Nucleic acids encoding polypeptide antigens from Chlamydia useful for
PT preventing, diagnosing and treating diseases such as community acquired
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
PT asthma -
XX
PS Claim 2; Fig 3; 174pp; English.
XX
XX
This sequence encodes a Chlamydia antigen of the invention, designated
CPN100635. The nucleic acids (and their complementary sequences) may be
used as diagnostic agents for detecting the presence of nucleic acids
encoding Chlamydia antigens in samples according to standard methods,
and therefore, for diagnosing Chlamydia infections. For example, they may
be used as primers and probes for diagnostic polymerase chain reaction
(PCR) assays. Antisense sequences may be used to down regulate
expression of the proteins and may be used to produce the protein antigens they
nucleic acids may also be used to produce the protein antigens they
encode according to standard recombinant DNA methodologies. The
proteins may then be used as antigens for the production of antibodies
(i.e. as vaccines) for preventing infection by Chlamydia. The
antibodies may also be used as diagnostic reagents for detecting
infections. Chlamydia is a pathogen implicated in the development of
CC (for example) community acquired pneumonia, upper respiratory tract
disease (especially bronchitis and sinusitis, asthmatic bronchitis,
adult-onset asthma and acute exacerbations of asthma in adults.
Sequence 2950 BP; 836 A; 593 C; 652 G; 869 T; 0 other;

```

PI Murdin AD, Oomen RP, Wang J;  
XX MPI: 2000-412339/35.  
DR P-PSDB: AAY90239.  
XX  
PT Nucleic acids encoding polypeptide antigens from Chlamydia useful for  
PT preventing, diagnosing and treating diseases such as community acquired  
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset  
PT asthma -  
XX  
PS Claim 2: Fig 5; 174pp: English.

XX This sequence encodes a Chlamydia antigen of the invention, designated  
CC CPN100638. The nucleic acids (and their complementary sequences) may be  
CC used as diagnostic agents for detecting the presence of nucleic acids  
CC encoding chlamydia antigens in samples according to standard methods,  
CC and therefore, for diagnosing chlamydia infections. For example, they may  
CC be used as primers and probes for diagnostic polymerase chain reaction  
CC (PCR) assays. Antisense sequences may be used to down regulate  
CC expression of the proteins and may be used to treat infections. The  
CC nucleic acids may also be used to produce the protein antigens they  
CC encode according to standard recombinant DNA methodologies. The  
CC proteins may then be used as antigens for the production of antibodies  
CC (i.e. as vaccines) for preventing infection by Chlamydia. The  
CC antibodies may also be used as diagnostic reagents for detecting  
CC infections. Chlamydia is a pathogen implicated in the development of  
CC (for example) community acquired pneumonia, upper respiratory tract  
CC disease (especially bronchitis and sinusitis), asthmatic bronchitis,  
CC adult-onset asthma and acute exacerbations of asthma in adults.  
XX  
SQ Sequence 2950 BP; 855 A; 724 C; 562 G; 809 T; 0 other;

Query Match 18.6%; Score 518.6; DB 21; Length 2950;  
Best Local Similarity 53.1%; Pred. No. 7.5e-141;  
Matches 1421; Conservative 0; Mismatches 1179; Indels 75; Gaps 12;

OY 143 AAGATGCTGAGTACCTATCTTAAAGGAAATGTCACTAGAAAATTTCTTGAA 202  
DB 258 atgcagatggaactatctatactaacaggagatgtctc--atcaccaatgcagat 314  
OY 203 CAGGACGACATACACAAAAGCTGTTTAAACACTAAGGCGATTGACTTTACAG 262  
DB 315 ctccgacgctcctaacgcctctcgtctttaaagaactactcgtggaactctctccaag 374  
OY 263 GTAACGGGAAGCTCTATTGTTTCCAAACGGTGTGACAGGACTGTAGCAGGGGCTGTC 322  
DB 375 gccacggctacccaattctcctcacaataatcgatgcggg-----agcgaactgtacct 428  
OY 323 TTAAACAGACCGTGTAGATTAATACACGTTTAAAGGTTTCTGCGCTATCTTTTA 382  
DB 429 ttacacaatcagctgaataaagctctcctccttccagattcctcctatcttgaactaa 488  
OY 383 TTGCGTCTCTGGAAGTTGATTAACACGCGAAAGAGCCGTTACTGCTTACGCGTA 442  
DB 489 tac-----aaaccacgaatgctacacacaggaacaggaacatcaagctccacagga 539  
OY 443 GCTTGAAGTTTGACAAAATGTGAGTTTGTCTGACGAAAACCTTTGAAGGATTAAG 502  
DB 540 ctgtctctatctcagtcgaactaagctgctactcttgcccaaaacttcttaagcaag 559  
OY 503 GCGGTCTATACCGCAAAAACCTTTTCAATTAACAGGACTACAAATGTGAGCTGTGTT 562  
DB 600 gaagcgccctccaagcagctatcagctatcgctaaacccaactaac---gtttg 656  
OY 563 CTGAATAATACCTCTCAAGAAAGGCGGACCATTTACAGCTTCGATGCCCTTACATTA 622  
DB 657 ccaaaaacaagaacgaacgaagggggtccctctattccacggaaggaattacaata 716  
OY 623 CTGGAACCAAGGAGAGTCTTTTCTGACAAATCTTCTGAGATTCGAGCTGCA 682  
DB 717 acataatcgttaactcagcatcttctcgaataatccgcggcgaacaatgagcgacca 776

OY 683 TTTTACAGAGCCGCTGCTGACTATTCTTAATAGCTAAAGTTCTTATTGACAAATA 742  
DB 777 tttaacaggaagctagcagttatttagcagcaacaagaacttagcttataacaata 836  
OY 743 AGGTACAGAGAGGAGCTCTCTCAACACGGGGATATGTGAGAGGTGCTATGTGCTT 802  
DB 837 gtlgagacg-----caacctcagctcaggggaggaatcattcagtagta 881  
OY 803 ATTAACCTAGTACAGATTAAGTAAAGTCCCTCATCGGAATTCAGATGTTCTCTACGA 862  
DB 882 gtacatacagccccaacacacagctcttaactctatcagaacacggggaactgaattatag 941  
OY 863 ACAATACATGACACACGCGGAGAGCTATCTATGTGAAGAAAGCTGAACGCTTCCG 922  
DB 942 gaatacagcaatcattagtagtggggcgattatactagaactcagttcttctctg 1001  
OY 923 GAGGACTTACCCCTTATTCAGTAAATAGTCAATGAGAGTACAGCTCTTAAAGGTGAG 982  
DB 1002 gagagctacgcttlltaaaaaaacactctgctatgatactgagctcccttagagagag 1061  
OY 983 CCATAGCTATCGAAGATGAGTGGGAATTTGATTCGCCCGATAGTGGTACATTGTCT 1042  
DB 1062 caatgcatctgctgactctgactcttcttgactcttcgactctgttgagacataact 1121  
OY 1043 TTTTGAAGATACAGT-----CACTTCTACTACTCTCTGGAGCAATVAGAGTATGA 1093  
DB 1122 ttgaaggaacacacagatgataagagctcttcctcgatgaaccattaccagaattcca 1181  
OY 1094 TCGACTTAGGA---ACGAGTGCMAAAGATGACACTTTGCTTGTGCTGCTGAGAGCCA 1150  
DB 1182 ttaacalcggaacacacacagctcgaagattgtacagctcgcagctccccaagcaatacta 1241  
OY 1151 TCTACTTCTATGATCCCATCTACTACAGGATCTTCCACACAGTTACAGATGCTTAAAG 1210  
DB 1242 tctactctatgctctctataacaactcagatcactgagctctcagatgctctaaact 1301  
OY 1211 TTAATGAGACTCCGCAAGATTCTGCACATATATACAGGGAACATCATCTTCAACAGAG 1270  
DB 1302 taatgctcctgcagctctgcaggaatccctcagatcaagaacacatcgtattcttcgag 1361  
OY 1271 AAAAGTTATCAGACAGACAGGCGCGAGATTCTTAAATCTTACTTGAAGCTACTACAGC 1330  
DB 1362 agaaagctctcgagaagagaagctgcaggaagctgataatcctcaactcaaatcagaac 1421  
OY 1331 CTGTAATCTTTTACGAGGACTACTTCTTAAACATGAGGAGTCTGACAGCTCAG 1390  
DB 1422 ccttaactcttgaggaggaagcaactctctcttaatacaggaagctcagttggtcgaagt 1481  
OY 1391 CATTTACTCAACAGGACAGATTTCTGTCGAAATGACAGTACACTCTAGAACTG 1450  
DB 1482 cctttcgaactctccgggctcaccctcctcctcagatgaggaacacatagaaacg 1541  
OY 1451 CTGATACTAGCACCATTAACAAATTTGGTCAATTAACTCACTGTATAGACGGTGCMAAG 1510  
DB 1542 ctgaatgagctactatcaatcaatcttctcctcaatgtagatcttcttaagaagacaaga 1601  
OY 1511 AGGCAAAAATAGAAACCAAGCTACGTCAAAATAATGCTTATCTATCTGGAACCATCTT 1570  
DB 1602 agggacgcttaaaagaacaacaagaacagcagcagcagcattatcttgatgcctctc 1661  
OY 1571 TATTGACCCGAGCGGACGCTTTATGAANAATCATAGTTTAAAGAAATCCTCACTAGC 1630  
DB 1662 ttgtagatccttcttggaatgltcagaagatgtctcttggaataaacctcaggtctt 1721  
OY 1631 ACATCTTAGAGCTCAAGCTT-----CTGGAATCTGTAAACACCGGACGTAGCTCCAG 1684  
DB 1722 ctgtctcactctactgctgacgaccccgagaaatltacacacagaccttagctgcg 1781  
OY 1685 ATCTTAAATGGGTGAGAAATTCATTAAGGCTATACGGAATCTTGAGGCCCAATTTGTT 1744  
DB 1782 atcccttagaaaaaatcctatccatcttggtgatacaagaaggaattgagcatatcttgc 1841  
OY 1745 GGGGACACGGGGC-----TTCTAGCACTGCAACCTTCAACTGAGCTAAAACTGGCTATA 1798

OY	1216	GAGATCTCGGCGAGATTTCGCGCTACAAATTTATACAGGGGAACCAATCATCTTTCGACAGAGAAAAG	12715
Db	1235	aaagctgcgcgaagatctgtaagaataacttaagaagctgaactcgtttcccttcggaagaaa	1354
OY	1276	TTATACAGACAGAGGCCCGAGATTCTTAAAAATCTTACTTCGAAAGCTACTACAGCGCTGT	1335
Db	1355	ctctcaagaagagaacttaagaagaacctgcacaactctgaagcttcaacttaccagaagcctgta	1414
OY	1336	ACTGTTTACGAGGATCTACTCATCTTTTAAACATAGAGTGACACTCTGACACTCAGGACTTC	1399
Db	1415	gagctctgcgaagctgcctcctaagatctgaaagaatgcagtgacgtgtagtctgcaataactata	1474
OY	1396	ACTCAACAGGCGAGATTTCGCTCTCGAAATATGACAGTACAGAACTACTCTAA---ACCTCT	1452
Db	1475	acgcaggtcgcagagatctgaaagtcgttatgatatgcaggggagactacttctgagcaagcgct	1534
OY	1453	GATCTAGCACACCATTAACCAATTTTGCTATTACATCATGTTCTATAGACGCTGCAAGAAG	1512
Db	1535	gaggggagatcactccaatgagctgactgacccaataatagatctcttagatctgagcaataata	1594
OY	1513	GCAAAAATAGAAACCAAAAGCTACGTCAAAAAATCTGACTTTATGGAACCACTGACTTTA	1572
Db	1595	gctatcattaaagcgagacggcagcaagtaagatggttccttaccagggcctatcgtct	1654
OY	1573	TTGGACCCGACGGGACGTTTATGAAATATCATAGTTTAAGAATTCGATCGCTACGAC	1632
Db	1655	gtagatgtcgcagggaaacttatgagcatcataactcagtcacaagaagctcttccct	1714
OY	1633	ATCTTAGAGCTCAAAAGCTTCTGGAACGTATACAGAACCGGACGAGTACTCTCAATCTCTATA	1692
Db	1715	ctaaatgagctcttcgcacaagaagaatgactactacagatalcccgataccccaat	1774
OY	1693	ATGGGTGAGAAATTCATTAACGGCTATACAGGAATCTTG-----GGCCCAATTTGTTGG	1746
Db	1775	ctaatactacgaataactatggtatgcaaggaaactgataatggttggtcgaagat	1834
OY	1747	GGGACAGGGGCTTACGACTGCAACCTTCACTGACTTAAACCTGGCTATATTTCTTAAT	1806
Db	1835	gcaactgnaaaaaaataaataatgctactcttaactcttgacataaacaagataatacgaagcaat	1894
OY	1807	CCCAGCGTATCGGCTCTTATAGTCCCAATATAGCTTATGGAATGCAATTTATATTTATG	1866
Db	1895	ccagaaagctcaggaactctggttccctaataagcgtcttgggtcttcttgcgaatgcgc	1954
OY	1867	TCCTCTCATTTCTTATGAGACTGCAAAAGAGGGTTCAGGAGACGTCGCTTTTGG	1926
Db	1955	tcacattcagagccctacatgcagcgagcaagtcgttactcttgcacaacaatttgcg	2014
OY	1927	TGTCGTGGATTATCTACTCTCTCCCTTAAGATAGTACAAAAACACGACGGCGGTTTGC	1986
Db	2015	gtatcaggaacgcggagactcttgcataagaaatcagaagaagaaacaaacgaatgattcgt	2074
OY	1987	CATTGAGATGCGGTATGTCATATAGAGGAAACCTCATACTGTTGTCAGATATGATTCCT	2046
Db	2075	cattcagcgcggttatgattagaagagagatcttcaagcgtcttgcgaattctct	2134
OY	2047	AGTCTGCAATTTTGTACGCTCTTGGAGAGATAGAGACTACTTTGTAGCTAAGATCA	2106
Db	2135	aatttctcttttgcagctcttcttgcctacgaagaagacacatctgtgcttaagaacat	2194
OY	2107	GGTACAGCTACGAGAGAACTCTCTATTACAGCAACAAAGAAACCTATATCTCTCTCT	2166
Db	2195	accacatgataatgcagggcgaatgcagttacgcacacccctgcagagcttaagaacctgcct	2254
OY	2167	TGCAAACTACGGCCCTGTTGCTGTATTATGTTCTTACAGAGATTCTGTTCTCTTTCA	2226
Db	2255	aaga-----tttgcagaagaattcgtacctcctactcttgcctcaat	2299
OY	2227	GGAACCTTATGCTACCCCATATGAGATACGATACGAAACCAATATATCAACATATCT	2286
Db	2300	gtctcggttcttatgcctatccgacaataatgcacaagaatcaactgcgtctatctt	2359

[illegible]





```

Db 2037 tgaaaacttaattagcttgccttcttgcgaactcttcttgtaagcgaataagcttctagt 2096
Oy 2094 acgttaagatcgaagctacagctacgaggaacgtctatttccgacgaacgaacgta 2153
Db 2097 cgtctaatactacacgatactatgcagagcgtctctatatacccaacacatacgaag 2156
Oy 2154 tttctctcttcttcccaactacgagcgttctgcttcttcttcttcttcttcttcttctt 2213
Db 2157 tagtgggtctcataggttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2213
Oy 2214 tcttctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2273
Db 2214 cctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2273
Oy 2274 tttcaacatattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2333
Db 2274 tttctctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2333
Oy 2334 tttcaacatattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2393
Db 2334 agctctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2393
Oy 2394 gaaatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2453
Db 2394 caaactgaatctgacatactatagcagcagcttcttcttcttcttcttcttcttcttctt 2453
Oy 2454 tttcaacatattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2513
Db 2454 atcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2513
Oy 2514 gaaatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2573
Db 2514 gttctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2573
Oy 2574 tttcaacatattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2633
Db 2574 cgtcaatgaatctgacatactatagcagcagcttcttcttcttcttcttcttcttcttctt 2633
Oy 2634 tttcaacatattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2693
Db 2634 tttcaacatattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2693
Oy 2694 tttcaacatattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2753
Db 2694 tttcaacatattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2753
Oy 2754 tttcaacatattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2786
Db 2754 tttcaacatattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2786

```

```

RESULT 6
AAK06821
ID AAK06821 standard; DNA: 2757 BP.
XX
AC AAK06821;
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp9 DNA.
XX
KW Omp9; outer membrane protein 9; surface exposed protein; antigen;
XX infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
OS Chlamydia pneumoniae.
XX
XX W0968953-2.
XX
XX 30-DEC-1998.
XX
XX 19-JUN-1998; 98WO-DK00266.
XX
XX 23-JUN-1997; 97DK-0000744.
PR

```

```

XX (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.
XX
XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mydind P;
XX
DR WPI: 1999-105610/09.
P-PSDB: AAM06822.
XX
XX Species-specific test for identifying mammals infected with
XX Chlamydia pneumoniae - comprises detecting antibodies specific for
XX outer membrane proteins of C. pneumoniae or nucleic acids encoding
XX these proteins
XX
XX Claim 6; Page 55-56; 115pp; English.
XX
XX This DNA sequence codes for the novel 96.7 kDa surface exposed
XX protein Omp9 (see AAM88422) of the human respiratory pathogen
XX Chlamydia pneumoniae. By generating antibodies against C.
XX pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
XX was obtained which reacted with outer membrane proteins. The
XX antibody was used to identify the genes (see AAM8816-27) encoding
XX Omp4-Omp15 proteins (see AAM88417-28) in an expression library of
XX C. pneumoniae DNA. The genes are situated in 2 gene clusters:
XX Omp12, 11, 10, 5, 4, 13 and 14 in one cluster and Omp6, 7, 8, 9 and 15 in
XX the other, and encode polypeptides of about 89.6-100.3 kDa and
XX about 56.1 kDa. The invention provides a new species specific test
XX for identifying mammals (including humans) infected with Chlamydia
XX pneumoniae. The test comprises detecting antibodies specific for
XX Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
XX membrane proteins, especially by PCR. The proteins are also used
XX in the diagnosis of C. pneumoniae infection in mammals. The
XX nucleic acids and proteins can also be used in the immunization of
XX mammals, the nucleic acids being particularly useful as DNA
XX vaccines for eliciting in vivo expression of antigens. The
XX vaccines may also prevent atherosclerosis and bronchial asthma,
XX which are possibly associated with C. pneumoniae.
XX
XX Sequence 2757 BP; 770 A; 564 C; 642 G; 781 T; 0 other.
XX

```

```

Query Match 19.0%; Score 528.6; DB 20; Length 2757;
Best Local Similarity 53.5%; Pred. No. 8.8e-144;
Matches 1428; Conservative 0; Mismatches 1149; Indels 90; Gaps 12;
Oy 142 GAAAGTCTGCACTACTACTATTTAAAGCAATGTCCTCTGAAATTTCTCTGCA 201
Db 142 GAAAGTCTGCACTACTACTATTTAAAGCAATGTCCTCTGAAATTTCTCTGCA 201
Oy 160 gatgtatgtgacgacatacttctgattggtgattctcgata--agcaagcaggg 216
Db 160 gatgtatgtgacgacatacttctgattggtgattctcgata--agcaagcaggg 216
Oy 202 ACAGGACAGCAATCAAAAGCTGTTTAAACACTAAGGGGATTTGACTTTCACA 261
Db 202 ACAGGACAGCAATCAAAAGCTGTTTAAACACTAAGGGGATTTGACTTTCACA 261
Oy 217 aacaaacagacttaacaaagtgttcttctaacacgcaggaatcttctcttcta 276
Db 217 aacaaacagacttaacaaagtgttcttctaacacgcaggaatcttctcttcta 276
Oy 262 GGTAAAGGAACTCTCTATTTGTTCCAAACGATGACAGGAGCTGTACAGGGCTGCT 321
Db 262 GGTAAAGGAACTCTCTATTTGTTCCAAACGATGACAGGAGCTGTACAGGGCTGCT 321
Oy 277 gggaaagcagatttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 336
Db 277 gggaaagcagatttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 336
Oy 322 GTTAAAGCAGCGGTGTAGATTAATCTACACGTTTATGSGTTTCTTCCGTACTTTT 381
Db 322 GTTAAAGCAGCGGTGTAGATTAATCTACACGTTTATGSGTTTCTTCCGTACTTTT 381
Oy 337 gttacatacagcagcttcttcttcttcttcttcttcttcttcttcttcttcttctt 396
Db 337 gttacatacagcagcttcttcttcttcttcttcttcttcttcttcttcttcttctt 396
Oy 382 ATTGGGTCTCTCGAAGTTGATTAACCGCAAGGACCGTTAGTGTCTACAGGGT 441
Db 382 ATTGGGTCTCTCGAAGTTGATTAACCGCAAGGACCGTTAGTGTCTACAGGGT 441
Oy 397 ctgtcagcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 444
Db 397 ctgtcagcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 444
Oy 442 AGCTTGAAGTTTGACAAAATGTCAGTTTCTTACGAAAACCTTTCAAGGATAAT 501
Db 442 AGCTTGAAGTTTGACAAAATGTCAGTTTCTTACGAAAACCTTTCAAGGATAAT 501
Oy 445 gttctgtgtttagagataggaatcttgcaaaaatgaaatgcttcttcttcttctt 504
Db 445 gttctgtgtttagagataggaatcttgcaaaaatgaaatgcttcttcttcttctt 504
Oy 502 GCGGCTGTATACCGCAAAACCTTTTCAATACAGGAGACATGACATGCTGTGTT 561
Db 502 GCGGCTGTATACCGCAAAACCTTTTCAATACAGGAGACATGACATGCTGTGTT 561
Oy 505 ggggagacatacagaagacttcttcttcttcttcttcttcttcttcttcttcttctt 564
Db 505 ggggagacatacagaagacttcttcttcttcttcttcttcttcttcttcttcttctt 564

```



CC omp4-Omp15 or detecting nucleic acid fragments encoding these outer  
CC membrane proteins, especially by PCR. The proteins are also used  
CC in the diagnosis of C. pneumoniae infection in mammals. The  
CC nucleic acids and proteins can also be used in the immunization of  
CC mammals, the nucleic acids being particularly useful as DNA  
CC vaccines for effecting in vivo expression of antigens. The  
CC vaccines may also prevent atherosclerosis and bronchial asthma,  
CC which are possibly associated with C. pneumoniae.  
XX  
SQ Sequence 2815 BP; 790 A; 575 C; 626 G; 824 T; 0 other;

Query Match 19.2%; Score 535; DB 20; Length 2815;

Best Local Similarity 53.0%; Pred. No. 1,2e-145; Mismatches 1180; Indels 75; Gaps 10;

147 TGTGTGAACCTACCTATTTAAGGAATGTCACTCTAGAAAATATCTCGAAGCAGG 206  
156 tactgataagactactctgacagagatatactctgcaaaacctgggagctcgg- 214  
207 CACGCAATCCAAAAGCTGTTTAAACAACATGAAGGCGATTTGACTTCACAGCTAA 266  
215 --cagcttaacgaagggtgttcttgcacactaagcgttaagcttgcgcgtaaa 272  
267 CGGGAACCTCTCTATTTGTTCCAAAGGTGATGACGAGCTAGACGAGGCTGCTTAA 326  
273 ggggtactactcttcttctttaaataatagctagctgctgaagcgagactcttctgt 332  
327 CAGCAGGTGTGATTAATATCTACACGCTTAAAGGTTTCTTGGCTATCTTTAATTC 386  
333 tacaactgataaataatcgtctgctaacaagatttcgaagcttacttcttagcgccc 392  
387 GTCGTCGGAAGTTGATTAACCTACCGCAAGAGCGCTTGCGTCTACGAGTACGTT 446  
393 atctcgtgtaacacacccccccaagaaagtgacgttaaaagtgaggggagtc---t 449  
447 GAATTTGACAAAATGATGATGCTGCTTTCAGCAAAAATTTTCAACGAGTAATGCGG 506  
450 taacttggataacaatggaactatcttataaacaagaattactcgtgaggaataatg 509  
507 TGTATACCGCAAAAATCTTTCATTACACGAGCTACAAATGTCAGCTGTTTCTGA 566  
510 agcaattcttcaacaagaatcttcttgaaaaacagcagcgatcttcttgaagg 569  
567 AAAA-----CCTCCTAAGAAAAGCGGAGCCATTACAGCTTCGATGCCCTTAC 617  
570 gaaataaatcgagcgcaacaaggaataaagtgggctatttgcctactgtaagctaga 629  
618 CATTACTGAAAACCAAGGGAAGTCTCTTTTTCGACAAATATCTTCGATTCGAGC 677  
630 tattacaataataatcgctctctactctctctgcacaataatctgtaagctgaggtg 689  
678 TGCATTTTTCACAGAGCCCTCGGTGACTATTCTAATATGCTAAAGTTCTTTAATGA 737  
690 agcataataatgacaggaactctgacattacagaggaatcgtctcttgaatttctcga 749  
738 CAATTAAGGTACAGAGAGCTGCTCCTCAACAAGGGAATATGACAGAGTGTACTCTG 797  
750 aaaaagtgtgac--agcgacgcgagaaatgtagagctcttctcgtgagatgcg---- 802  
798 TGCTTATTAACCTAGTACGATACGATACGATACGATACGATACGATACGATACG 857  
803 -----atgttaccatactcgggagtaagtaagctgtaacttt 836  
858 CAGCAACATATCATCGAACAACGCGGAGAGCTATCTATGTGAAAAAGCTCGAATGCG 917  
837 ctcaaggaaccaaagctgtagctaatgycgagcattatgctaagaagcttaccatgcgc 896  
918 TTCGGAGAGACT-----TACCTATTCAGTAAATATAGTCAATGAGAGTACAGCTCC 971  
897 ttccgggggggggggggggtatctctcttcttcaacaatagcttccaaagctaccatcgag 956  
972 TAAAGGTGAGCACTAGCTATCGAAGATAGTGGGAATGAGTTATTCGCGCATAGTGG 1031

957 taatggtgagaccattctatactcagcagctgagagagtgagcttctcagcagaagcgag 1016  
1032 TGACATTTGCTTTTATGGAATPACAGTACTTCTACTCTC---TGGAGCATATGAAAG 1088  
1017 ggaattctcctcaatggaatgcatgtctgtcgaactcaccacaataaataaagaa 1076  
1089 TAGATGACTTAGAAGAGAGTGCAGAAAGATGACAGCTTGGCTTGTGCTGGTAGAGC 1148  
1077 ttctatgacataaggaattactcgaagaagatcgaattactagtgataatcttggcagatag 1136  
1149 CATCTACTCTATGATCCCATTAACATACAGATCTTCCAAACAGTACAGATGCTTAA 1208  
1137 catcttcttcaagatccgatctactgctaataagctcgcgagcttctacagctacttaaa 1196  
1209 AGTTATATGACTCCGCAAGATTCTGCACTACATATACAGGAACATCATCTTACAGG 1268  
1197 tctcaataaagctgtgatgaggaatagatagatataagtgagtgatgttcttctg 1236  
1269 AGAAAGTTATCAGACAGAGCGCGAGATTCTAAATCTTATGGAAGCTACTAC 1328  
1257 tgaagaagctcctgaaagatgaaagcaaaagttgcagacaactcactcactcagcgaagca 1316  
1329 GCTGTACTCTTTCAGAGAGTACTCTATCTTAAACATGAGTACGATCTGACAGCTCA 1388  
1317 gctgtactacttaactgaggaatattagtaacttaaaagtggtgactcctcgatagaa 1376  
1389 GGCATTCACCTAACAGCAGAGATTCTCGTCTGAAATGAGAGTACGACTACTTGAAC 1448  
1377 aggttctactcagaacgggggtctctctgtatattatgagtcgagcaaaagcttaaaagc 1436  
1449 --TGCTGATCTACTGACACCATTAACAAATTTGTCATTAACTCAGTTCTATAGAGCGTGC 1505  
1437 aagtaacagaagaggtacactttaaagaagcttcttcaacttcttgaactctttagcgaggg 1456  
1506 AAAGAGGCAAAAATAGAAACCAAGCAATGCTCAAAAATCTGACTTATCTGGAACCAT 1565  
1497 taagaagtgtaattgtctgtctctgcagcaagtaaaatgtagcccttagtgctcgat 1556  
1566 CACTTTATGACCCGAGCGGACGCTTTTATGAAATATCATAGTTAAGAAATCCTCAGTC 1625  
1557 tctcttcttgataccaagaaggaatgcttgaataatcagactaggaataaactcaaga 1616  
1626 CTAGACATCTTGAAGCTCAAGCTTCTGAACTGTAAACAGCAACCGGAGTACCTCAGA 1685  
1617 ctcttcaattgtcgaactcctctgtctggtgacgcgaacactcagatgttccagcggt 1676  
1686 TCTATATATGGGTAGAAATTCATTACGCTATCAGGAACTTGGGCCCCAATTTGTTG 1745  
1677 tctcaagtagcaactcctcactagcaactatggtatcaaggtacttggggaatgacttgggt 1736  
1746 GG-----GACAGAGGGGCTTCTAGACACTGCAACCTTCAACTGAGATAAATCG 1793  
1737 tgaatataccgcaagcactccaagaactcaagacagcagacttgaacttggacaatacag 1796  
1794 CTATATTCCTATCCCGAGCTATCCGCTTATGCTTCTTATGCTTATGGAATGCAAT 1853  
1797 ctacacttcgaatccctgaggtcgaagcaagccttgaacttctcaatgacttgggagatctt 1856  
1854 TATATATATAGCTCTCTCCATTTATCTTATGAGAGTCAACAGAGGTTGACGAGAGA 1913  
1857 ttcaagataccaagcagatcaagtgatcatagagaagtgcttggacttcttgcaga 1916  
1914 CCGTGTCTTTTGGTGTGCTGATATATCTACTCTTCCATTAAGATATGTAACAAAACAG 1973  
1917 tcaagagcttccggtcgtcgagagtcgccaatttcttgaataaagaataaagagggaaa 1976  
1974 ACGGSGGTTTCGCATTTGAGTGGCGGTATGTCATAGGGAACCTACATATCTTGTTC 2033  
1977 acgcaataatccgataataatcgttgatagatagctacgaggggtgcagcgcaactgtctc 2036  
2034 AGATTAATTTCTTATGCTGCTGATTTTGTACGCTTGTGGAAGAGATAGAGACTACTTGT 2093



DT 27-FEB-2001 (first entry)  
 XX Chlamydia pneumoniae genome DNA.  
 DE Chlamydia pneumoniae genome DNA.  
 XX Genome; diagnosis; vaccine; ds.  
 KW Chlamydia pneumoniae.  
 OS Chlamydia pneumoniae.  
 XX MO200027994-A2.  
 XX 18-MAY-2000.  
 PD 12-NOV-1999: 99MO-US26923.  
 XX 12-NOV-1999: 99MO-US26923.  
 XX 12-NOV-1999: 98US-0108279.  
 PR 08-APR-1999: 99US-0128606.  
 XX (REGC ) UNIV CALIFORNIA.  
 PA Stephens R, Mitchell W, Kalman S, Davis R:  
 PI WPI: 2000-376516/32.  
 DR WPI: 2000-376516/32.  
 XX Isolated nucleic acid for use in diagnostic and analytical methods  
 PT encoded genomic sequence of Chlamydia pneumoniae -  
 PS Claim 2: Page 128-320: 320pp: English.

CC This invention describes a novel nucleic acid (N1) encoding a Chlamydia  
 CC pneumoniae protein (PI), given in the specification. The isolated nucleic  
 CC acid is useful for diagnostic and analytical methods, such as,  
 CC hybridization-based assays or amplification-based assays. The protein may  
 CC be used for diagnostic purposes, for their enzymatic or structural  
 CC activity, or as a vaccine. The invention also describes (1) a probe  
 CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid  
 CC (N2) that hybridizes under stringent conditions to N1; (3) an expression  
 CC cassette comprising N1 under the transcriptional regulation of a  
 CC transcriptional initiation region functional in an expression host, and a  
 CC transcriptional termination region; (4) a cell comprising an expression  
 CC cassette of (3) as part of an extrachromosomal element or integrated into  
 CC the genome of a host cell as a result of induction of the expression  
 CC cassette into the host cell, and the cellular progeny of the host cell;  
 CC (5) a method for producing a PI comprising growing a cell of (4) where  
 CC the protein is expressed and isolating the protein free of other  
 CC proteins; (6) a purified polypeptide composition comprising at least 50  
 CC weight % of PI; and (7) a monoclonal antibody binding specifically to the  
 CC peptide of (6).  
 CC  
 XX Sequence 273254 BP: 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

Query Match 19.2%; Score 536; DB 21; Length 273254;  
 Best Local Similarity 53.1%; Pred. No. 6.7e-145;  
 Matches 1418; Conservative 0; Mismatches 1180; Indels 74; Gaps 10;

QY 147 TGTGTGAACACTACTTATTTAAGGAATGACACCTGGAATAATTCCTGGAACAG 206  
 Db 98349 Tactygaataagactaactaactcgaagataataactcgcgaacccctgggattcgcg- 98407  
 QY 207 CACAGCAATCAGAAAAAGCTTTTAAACACACTAAGGGGATTTGACTTTCACAGTTAA 266  
 Db 98408 --cagcttaacgaaggtgtttcttcgactaagcgaactcttaagcttcgcgtaa 98465  
 QY 267 CGGGAACCTCTTATTTCCTCAACAGCGTATGAGGAGCTGTGACAGGGGCTCTGTTAA 326  
 Db 98466 Tactygaataagactaactccttttaataataagctagtcggaagcgacgacattctcgt 98525  
 QY 327 CAGCAGCGTGTGATTAATCTACACGTTTATAGGTTTCTTCCTATCTTTAATGTC 386  
 Db 98526 Tactaactgataaaatcctcgtcgaacgaagattcgaagcttaactcttagcggccccc 98585  
 QY 387 GTCCTCTGGAAGTTGATTAACCTACCGCAAGAGGCGCTTAGCTGCTACGAGGTAGCTT 446  
 Db 98585 Tactaactgataaaatcctcgtcgaacgaagattcgaagcttaactcttagcggccccc 98585

Db 98586 atcatcgtlaatacaaacccccctcgaagaaagtgcaagtaatgtaggaggtac--t 98642  
 QY 447 GAAGTTGGACAAAAAATGCTAGTTGCTCTCAGCAAAAACCTTTTCAACGGATATAGCGCG 506  
 Db 98643 Tacttggataaacaatggaactattttaaacaagaattactcgtggaagaatgtagcgg 98702  
 QY 507 TGCATACCCGCAAAAACCTTTTCAATTAACAGGAGACTACATGACGCTGCTTTTCTGA 566  
 Db 98703 agccattctaccagaatcttctcttgtaaaaacagacggatcgatcttcttgaag 98762  
 QY 567 AAATA-----CCTCTCAAAAAGAAAGCGGAGCCATTCAAGCTTCGATGCCCTTAC 617  
 Db 98763 gaataaactcgcgcgaacgaagaaaggtggggcatttgcctactgtagtactgaga 98822  
 QY 618 CATTAAGTGAACCAAGGGGAAGTCTCTTTTTCGCAATCTCTTCGATTCGTGAGC 677  
 Db 98823 Tattacaataataacgctctcctcctcctcctcgaacaaatgctggaagctgcgaagtg 98882  
 QY 678 TGCATTTTACAGAAAGCCCTCGGTGACTATTTCTAAATATGCTAAAGTTTCTTTATTTGA 737  
 Db 98883 agctataaataagcaacggaactcgttaacattacaggaataagctctctgattctcga 98942  
 QY 738 CAATTAAGTCAAGAGGAGCGAGCTCTCAACACGCGGAGATATGACAGAGGTCTATCTG 797  
 Db 98943 aaatagtgtagc---agcgacgcgagaatggaagagctcttcgtgagatgcg---- 98995  
 QY 798 TGCTTATTAACCTAGTACATATTAAGCTCACCTTCAGGAAATTCAGATGTTACTCTT 857  
 Db 98996 -----atgtaccatactcgggaatcagaatgtagtactt 99029  
 QY 858 CAGCAACATATACATTCGACACAGCGGAGGAGCTATCTATGTAATAAAGCTCAACATGCG 917  
 Db 99030 ctaagaaacaacgaagctgtagtgaatgtagcgaagcattatgcaagaagcttactcgc 99089  
 QY 918 TTCGAGAGACT-----TACCTATTCAGTAAATATGCTCAATGAGAGGTACAGCTCT 972  
 Db 99090 ttcgcgggggggggggggtatctctcttcttaacaataagtcacaaggtaccacgtcga 99149  
 QY 973 AAAGGTGAGGCCATAGTATCGAAGATAGTGGGGAATGCTTATCCGCGGATAGTGGT 1032  
 Db 99150 aatcgtggaagcattcttaactcgaactgcgcgcgagagtgtagtcttccaagaagaacga 99209  
 QY 1033 GACATGCTCTTTTATGAGGAGATACAGTACACTTCTACACTCTC---TGGGCAATTAAGT 1089  
 Db 99210 gaattactccttaagtggaatgtagtcatgtgtgcaactacaccacaactaacaagaat 99269  
 QY 1090 AGTATGACTTTAGCAAGAGAGTCAAAAGATGACAGCTTTGCGTTCTGCTGCTAGACCC 1149  
 Db 99270 tctattgacataagatctactcgaagaatcagaatttactgcaatctcgtgagcatagc 99329  
 QY 1150 AACTACTCTATGATCCCATTAATACAGAGATCTTCCACAACATTTACAGATGCTTAA 1209  
 Db 99330 atcttcttccaagatccgactactcgaactaagaacgctgcgagattctcagaactttaa 99389  
 QY 1210 GTTAATGAGACTCCGAGATTTCTGACTTCAATATACAGGGAACATCTCTCACAGCA 1269  
 Db 99390 ctcaataaagctcgtatgcgaagtaatagatataagtgagcgatttcttctgtc 99449  
 QY 1270 GAAAAGTTATCAGAGACAGAGCGCGAGATTTCTAAATCTTACTGCAAGCTACTACAG 1329  
 Db 99450 gaaaagctcctcgaagaatggaagaatgtagcaaacctccttaacgctcgaagcag 99509  
 QY 1330 CCTGTAACTCTTTTACAGAGTACTATCTTTAAACATGAGATGACTCTGCAAGACTCAG 1389  
 Db 99510 cctgtaacttaactcgcgaagaaattagtaactaaacgtagtgcactcctcgaactaga 99569  
 QY 1390 GCATTAATCTCAACAGAGAGATCTCTGCAAAATGAGAGTACACTACTAGAAC- 1448  
 Db 99570 gacttaactcagaacgcgggttctcctgtatataatgtagcgggacaaagcttaaaagca 99629  
 QY 1449 --TGTGATTAAGTCAACCAATTAATGCTTATTAACATGATGCTTATAGAGGCTGCA 1506  
 Db 99630 agtacagagagaggtcacttaacaggtcttccattccctcgttagactcttgaagcgaggt 99689

|||||  
Db 32477 aaagttccttattcgaaataaggtcaacagagcgagctccctcaacaacgggagatag 32536  
QY 781 TCAGAGAGGTGCTATCTGTCTTATAAAGTACTAGAGATACTTAAGGTCACTCCACTGGA 840  
Db 32537 tcaggaaggtgctatctgtgcttataaactagtacagataaggtcaccctcctgga 32596  
QY 841 AATCAGATTTCTCTCTGCAACAATACATGAGACAACGGGGAGGAGCTTCTATGTG 900  
Db 32597 aatcagatgttactcttcacgaacaatacgcacaacagcgaggaagctatcatgtg 32656  
QY 901 AAAAGCTCGAACTGCTGCTCCGAGAGACTTACCTATTAGTAAGAAATAGTGTCAATGA 960  
Db 32657 aaaaagctcgaactggtcttcgagagacttaccctatctcagtaagaatgttcaatgga 32716  
QY 961 GGTACAGCTCCCTAAAGGTGGAGCCATAGCTATCGAAAGATAGTGGGAATGTGATTATCC 1020  
Db 32717 ggtcagagctccctaaaggtggaagccatagctacgaagatggtggaatgtgattatcc 32776  
QY 1021 GCCGATAGTGTGACATTTGTCTTTTAGGGAATACAGTCATCTTCTACTCTCTGGAGC 1080  
Db 32777 gccgatagtgtgagacattgtctttagggaaatacagttacttactctctcgtggagc 32836  
QY 1081 AATPAGAATGATATGACTTGTAGGAAAGAGTGCAGAAAGATGACAGCTTTGGCTTCTGCT 1140  
Db 32837 aatpagaatgatactgactgacttaggaacaggtgcagaatgtcagcttgcgtctcgtct 32896  
QY 1141 GGTAGAGCCATCTACTCTTCTATGATCCCATPACTACAGATCTTCCACACATTTACAGAT 1200  
Db 32897 ggttagagccatctactcttcttctcccaataactagatctacacaaagcttcaagtt 32956  
QY 1201 GCTCTTAAAGTTAATGAGACTCCGGCAGATTTGTGACTTCATATATACAGGAAACATCATC 1260  
Db 32957 gctcttaaaagttaatgagactcgcgcagacttctgcactacaataatacgggaaactatcc 33016  
QY 1261 TTCACAGAGAAAGTTATCAGAGACAGAGGCCGAGATTTCTAAAAATCTTACTTCGAAG 1320  
Db 33017 ttcacagagaaagttatcagagacagagcgcgagacttcttaaaatcttactctcgaag 33076  
QY 1321 CATACAGACCCGTGTAACCTTTTCAGAGGTACTCTATCTTTAAACATGAGACTGCTCG 1380  
Db 33077 caatacagaccgttaactctcttcaggaaggtactcttactttaaacaatggaatgtactcgt 33136  
QY 1381 CAGACTCAGGACTTCACTACAGAGCAGATTCGTCTCGAATGAGACGTAGAACACTACT 1440  
Db 33137 cagactcagagacttaactcaacaacagagattcgtctcgaatgagcgttagaactact 33196  
QY 1441 CTAGAACCTGCTGATACTTACGACCATAAACAATTTGGTCATTACATCGTTCTATAGAC 1500  
Db 33197 ctagaacctgctgatacttagcaccataaacaatttgttcatttaacatcagttcttaagac 33256  
QY 1501 GGTGCAAGAGCAAAAATAGAAACCAAAAGTACGTCAAAAATCTGACTTTATCTGGA 1560  
Db 33257 ggttgcagaagaagcaaaaaataagaaccaaagctacgltcaaaaaatcgtcttatactgga 33316  
QY 1561 ACCATCACTTTATTTGACCCGACGGGCAAGTTTATGAAAATCATAGTTTAAAGAAATCCT 1620  
Db 33317 accatcaacttatttgaaccgcggcacgttttattgaataatcatagtttaagaatcct 33376  
QY 1621 CAGTCTACGACATCTTAGAGCTCAAAGTTTGTGAACGTTAACAAGCCGCAAGTACT 1680  
Db 33377 cagttcctaagacatcttgaagcttcaagcttcttgaacgttaacaagccgcagtgact 33436  
QY 1681 CCAGATCCATATATGGGTGAGAAATTCATTAAGGCTATACAGGAAGTTGGGGCCCAAT 1740  
Db 33437 ccagatccataataatggtgtagaataatccatctacgtctacaggaacttggggcccaat 33496  
QY 1741 GTTTGGGGGACAGGGGCTTCTACGACTGCACCTTCACACTGGACTTAAACTGGCTATAT 1800  
Db 33497 gtttgggggacaggggcttctacgactgcaaccttcaactgcaaaaaatggtcatatt 33556  
QY 1801 CCTAATCCGAGCGATTCGGCTCTTATGCTCCCTAATAGCTTATGGAATGCATTTATAGAT 1860  
|||||

Db 33557 cctaaccgagcgtatcgctctttagtccctaataagctatatgaaatgcatltagat 33616  
QY 1861 ATTACTCTCTCCATTTACTTATGAGAGACTGCAACGAAAGGTTTCAGGAGAACCGTGT 1920  
Db 33617 attactctctccatctattcttagagactgcaacgaaggttgcagggaaacggtct 33676  
QY 1921 TTTTGTGTGCTGATTTATCTACTTCTTCCATAAGGATATGACAAAAACAGACGCGGG 1980  
Db 33677 ttttgtgtgctgattatctactcttctccataagtagtatacaaaaaacagaagcggg 33736  
QY 1981 TTTTGGCCATTTGATGTCGGGTTATGTCTATAGAGAGAAACCTACATCTTGTTCAGATTA 2040  
Db 33737 tttcgcattttagtggcggttattgtcattaggaaggaacctacatactgttccagataag 33796  
QY 2041 ATTCTTAGTGTGCAATTTTGTACGCTCTTGTGAACAGATAGAGACTCTTGTACTAG 2100  
Db 33797 attccttagtgcgtacttcttgcagctcttgcgaagagataagagactacttgcataag 33856  
QY 2101 AATC-AAGGTACAGTCTACGAGAGAACTCTTATTTACAGCAACGAAACCTATATCTC 2159  
Db 33857 aatcaaggtatacgtctacgaggaactctctattacacgaacaagaaactatattcc 33916  
QY 2160 TCTTCTTGCAAACTACGGCCCTTGTTCGTTCTTATGTTCTACAGAGATTCCGTGCT 2219  
Db 33917 tcttctctgcaactacagcccttgcgttgcgttattgttccataagagattccgttct 33976  
QY 2220 CTTTTCAGAAACCTTAGCTTACACCATACGATTAACGATGGAACCAACTATACAA 2279  
Db 33977 ctttctcagaagaaaccttagctacaccatacgaatagatagctcgaaaaccaatataaac 34036  
QY 2280 AATCTCTACTGTTTAAAGAGAGCTGGGGGAAATGATAGTTTCGTTTAAATTCGTGGA 2339  
Db 34037 atactactgtttaaggaagcttgggggaatgatagttgcgtttaagaattcgtggag 34096  
QY 2340 AGCTCCGATTTCTTATGATGAAGAGTCTTATTTGACAGCTACATGCCCTTATGAAAT 2399  
Db 34097 agctccgatttgccttagatgaaaggtctctatttgcaggtacatgccccttataaatt 34156  
QY 2400 GCAGTTTGTCTATGACATCAGAGAGTTTAAAGAACGAGGAGACAGAAAGCTCGGAAT 2459  
Db 34157 gcaatttctatctgcacatcaggaaggtttaaagaacgaaggaacgaagctcgtgaatt 34216  
QY 2460 TGGAGTAGCCGCTGTGTGTAATCTTGCTTACCTATCGGATCGATTTGATTAAGGAATC 2519  
Db 34217 tggaaagtagcgcgtcttgcgaattcttgccttactatcgggaltcgatttgaagaa 34276  
QY 2520 AGACTGCCAAGATGACAGCTTACATCTTAACTTGTGTATGATGATGATCTGTCTGATG 2579  
Db 34277 agactgcgaagatgcaaacgtataacttactcttgcgttatacttgcgtctgtctgtag 34336  
QY 2580 TAACCCGACTGTACGACAACTCGAATTAAGCGTGATCTTGGAAAACTTCGCTGATC 2639  
Db 34337 taaccgcgactgtatacgaacaacctgcgaattagcgtgtacttcttggaaaacttgcgtac 34396  
QY 2640 GAATTTGGCAAGACGCTTTATGCTCTTCGTCAGAGGAACATTTTGTCTTAACTCAAA 2699  
Db 34397 gaatttgcgaagaacagctttagtccctcgttcagggaaacatttgcgtttaaactcaaa 34456  
QY 2700 TTTTGAAGCTTTTACCAATTTTCTTTTGAATTTGGTGGGTGATCTCGAATTAAGAT 2759  
Db 34457 ttttgaagcctttagccaatttcttcttgaattgcgtgtgtatctcgaattatacaatgt 34516  
QY 2760 AGACTTAGAGCAAAATACAAATTTCTAA 2787  
Db 34517 agacttagagcaaaaataaccaattctaa 34544  
|||||

RESULT 4  
AAC81914  
ID AAC81914 standard; DNA: 273254 Bp.  
XX  
AC AAC81914:  
XX

```

Db 2261 ctctccgcaaacacgacgacctgctgcttctatgtctctacagagatccctgtctc 2320
QY 2221 TTTTCAGAGAAACCTTACGTACACCCATACGATATACGATCTGAAAACAGATATACACA 2280
Db 2321 ttctcaagaaaccttaacataccacacacgataacgatacctgaaacacagatacacaca 2380
QY 2281 TATCTCTCTGTATAGAGAGAGCTGGGGGAATGATAGTTTCGCTTATAGATTGGTGGAAGA 2340
Db 2381 tatctactctgttaagaagaagctggggaatgatagttctgctttagaattccggtgaaga 2440
QY 2341 GCTCCGATTTGCTTATAGTGAAGAGTCTCTATTGAGACAGTACATACCCCTTCATGAAATG 2400
Db 2441 gctccgattctgttagatgaagaagctctattttagcagatacagccctcatgaattg 2500
QY 2401 CAGTTTCTATGACATCAGACAGAGGATTATAGACAGAGAAAGAGCTCCGATATT 2460
Db 2501 cagtttctatgacacacacaggaaggtttaaagaacaggaagaagctcgtgaattt 2560
QY 2461 GGAAGTACCGCTCTTGTAATCTTGCTTACCTATCGGATCCGATTTGATAGGAATCA 2520
Db 2561 ggaagtagccgctctgtgaatcttgccttacctacacggtatccgattgataagaatca 2620
QY 2521 GACTGCCAAGATGACAGCTACATCTACTCTTGTTATACCTGTGATCTTGTCTAGT 2580
Db 2621 gactgccaaagatgaacagtaacatacctctgttactctgtgatactgtctgtaagt 2680
QY 2581 AACCCGACGTACAGACAAACACTGCGAATTAAGCGGTATCTTGGAAGAACCTTGCGTACG 2640
Db 2681 aacccgacgttagaacaacacgtgaaatlagcgtgattcttggaaacaccttgcgtacg 2740
QY 2641 AATTGGCAAGACAAAGCTTATAGTCTCTGTCGACAGGAACATTTTGTCTTAACTCAAT 2700
Db 2741 aattggcaagaacaaagctttagtctctgcgtgcgggaacaaatctttagtaactcaat 2800
QY 2701 TTTGAGACCTTTAGCCAAATTTCTTTGAAATGCGTGGTCATCTGCAATTCATATGTA 2760
Db 2801 ttggaacctttagccaattcttcttgaattgctgtgcatctcgcaattacatgta 2860
QY 2761 GACTTAGAGCAAAATACCAATTTCTAA 2787
Db 2861 gacttagagcaaaataccaattctaa 2887

```

## RESULT 3

AAK91990  
ID AAK91990 standard; DNA: 1230025 BP.

XX AAK91990;

DT 13-SEP-1999 (first entry)

DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope; ss.

OS Chlamydia pneumoniae.

XX WO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

XX 21-NOV-1997; 97FR-0014673.

XX (BEST ) GENSET.

XX Griffiths R;

```

DR WPI: 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Claim 1; Page 291-611; 1912pp; English.
XX
CC The present sequence represents the complete genome of Chlamydia
CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes
CC respiratory disease such as pneumonia and bronchitis and is thought
CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
CC encoded by the open reading frames of the C. pneumoniae genome (see
CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.
CC Vectors containing C. pneumoniae nucleotides sequences can also be
CC used as immunogenic compositions, especially where the vector directs
CC the expression of a neutralising epitope of C. pneumoniae.
XX
SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other:

```

Query Match 99.1%; Score 2760.8; DB 20; Length 1230025;  
Best Local Similarity 99.98; Pred. No. 0;  
Matches 2784; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

```

QY 1 ATGAAGTCTTCTTCCCAAGTTTGTATTTCTACATTTGCTATTTCCCTTGCTATG 60
Db 31758 atgaagctctcttccccaaagtgtgtatttctacattgtctattcccttgctatg 31817
QY 61 ATTGCTACCGAGACAGTTTGGATTCAGTGCAGTTTCGATGGGAATTAATAATGCTAAT 120
Db 31818 atgctacgagacagatttggatccaagtgcgagtttcgagtggaaataaaatgtaat 31877
QY 121 TTTTCAGTTGCTGAGATCAGGAAGATGCTGGAACTACCTACCTATTATTAAGGAAATGTC 180
Db 31878 ttctcagctcgtgagagtcggaagatgcggaactaccacctaacttaagggaaatgctc 31937
QY 181 ACTCTAGAAATATATCTCTGGAACAGCAGCAATCATCAAAAAGCTGTTTAACTCAACT 240
Db 31938 actctagaataatctctctggaacagcagcaatcacaataaagctgtttaaacaact 31997
QY 241 AAGGCGATTTGACTTTCACAGGTACGGGAACCTCTATATTGTTCAACACGGTGATGCA 300
Db 31998 aaggcgatttgaacttccacaggtacaggtacgtctctatttccaaacgtgtgatgca 32057
QY 301 GGGACTGTACAGGGGCGTGTGTAACAGACAGCGGTAGATTAATCTACAGATTATATA 360
Db 32058 gggactgtacagggcggtgtgtaacagacagcggtgtagataaattctacacgattata 32117
QY 361 GGGTTTCTTCGCTATCTTATATGCTCTCTGGAAGTTCGATTAACCTACCGCAAGGA 420
Db 32118 gggtttcttcgctatcttattatgtgctccctggaaagttcgataactacggcaaaaga 32177
QY 421 GCCGTTAGCTGCTCTACGGGTACCTGAGTTTGACAAAAAATGTCAGTTGCTCTTACG 480
Db 32178 gccgttagcgtcttaagggtagcttg-agttgacaataatgcaagtttgccttcagc 32236
QY 481 AAAAAGCTTTCACGGAATATGAGGGGTGCTATCAACGCAAAAACCTGTTTATTAACAGG 540
Db 32237 aaaaagcttccaaggaataagcggtgctataccacgcaaaaacctcttattcaacagg 32296
QY 541 ACTACATGTACAGCTCTGTTTCTGAAATACCTCTCAAGAAGAGCGAGGCATTTACG 600
Db 32297 actacaatgtcacgctctgttcttgaaataactctccaagaagaagcgagacattcag 32356
QY 601 ACTTCGAGCCCTTACCATTAATCTGCAAAACCAAGGGAGACTCTTTTCTGCAATACT 660
Db 32357 acttcgagcccttaccattactcgtgaaccaaagggaagctcttcttctcgaataact 32416
QY 661 TCTTCGATCTTGAGCTGCAATTTTACAGAGCCCTCGGTGACTATTTCTAATAATGCT 720
Db 32417 tcttcgattctgagctgcaatttcaacagaagccctcgtgactatttctaataatgct 32476
QY 721 AAGTTTCTTTATTTGACATTAAGTTCACAGAGCGAGCTCTCAACACGCGGAGTATG 780

```

```
OY 1 ATGAGTCTCTTCCCAAGTTTGATTTTCTACATTGCTATTTCCCTTGTCTATG 60
DB 101 atgaagctcttccccaagtctgtattcttctacattctgtctattcccttgcata 160
OY 61 ATTGCTACCGACAGTTTGGATTCAAGTCGAGTTTCGATGGCAATAAAAATGTTAT 120
DB 161 attgctaccgaaagtttggattccaagtgcagtttcgaatggaaataaattgta 220
OY 121 TTTTCAGTTCGAGAGCAGGAAGATGCTGAATCTACTACTATTTAAGGAAATGTC 180
DB 221 ttctcaagctcvgaaagcagaagaagcvgcgaactaccctacttaagaagaaatg 280
OY 181 ACTCTGAAAAATATTCCTGGAACAAGCAGCAATACAAAAAGCTGTTTAAACAAC 240
DB 281 actcagaanaatattcccggaacagcagcaacacaaagctgtttaaacaacact 340
OY 241 AAGGCGATTGACTTTTACAGATTAACGGGAATCTTATTTGTTCCAAACGGTGA 300
DB 341 aaggcgatttgaacttccacagtgaaacgtctctattgttccaaacggtgaa 400
OY 301 GGGACTGTAGCAGGGGCTGCTGTTAAACAGACCGGTGATTAATTCACACGTTT 360
DB 401 gggactcgaagcgggcgcgcgttaacacagcgcgtgaaataaactccaacg 460
OY 461 GGGTTCCTGCTATCTTTATTTGCTCTCGGAAATGATTAATTCACCGCAAGA 420
DB 461 gggcttcttcgtctctcttatttgcgtctcccgaaagttcgaataccacgg 520
OY 421 GCGGTACCTGCTACGCGGATGCTGAAGTTTGACAAAAATGTCAGTTGCTTCA 480
DB 521 gccgttaagctctcacaagcggatgagcttgacaataaattgtaagtttgc 580
OY 481 AAAAATTTTCAAGGATTAAGCGGTGCTATTCACGCAAAAACCTTTCAATTA 540
DB 581 aaaaacttcaaggaataaagcggcgtcacaacgcgaataaacttcaataca 640
OY 541 ACTACATGTCAGCTCTGTTTCTGAAAAATCCTCTCAAAAGAAAGCGGAGCCAT 600
DB 641 actacaatgtaagctctgttcttctgaataactctcctaagaagcgggacat 700
OY 601 ACTTCGATGCCCTTACCATTAATCTGAAACCAAGGGGAAGTCTTTTTCGACA 660
DB 701 acttcgattgccttaccattacttgaaccaaagggaagctcttcttctga 760
OY 661 TCTTCGATTCGAGCTGCAATTTTACAGAGCCTCGTGATTAATTTCTAATA 720
DB 761 tcttcgatttcggaagcgcgaatttctacagaagcctcgcgtgaaacttcta 820
OY 721 AAAGTTTCCTTATTTGACATTAAGGTCAAGAGCAGCTCCTCAACAACGGGGAT 780
DB 821 aaagtttccttattatgaataaagtcacaagcggagcgtcctccaacaa 880
OY 781 TCAGAGAGTGTCTATGTGCTTATTAACACTAGTACAGATTAAGGTCAACCT 840
DB 881 tcaaggaagtgctatctgtctataaactagtaacagatacctcaccctcag 940
OY 841 AATCGATGTTACTCTTGAACAATACATCGAACACGGGGAGGCTATCTATG 900
DB 941 aatcgaatgttaacttcaagaacaatacactgacaacagcggagagcctac 1000
OY 901 AAAAAGCTGCAAGTGGCTTCGAGAGACTTACCCCTATTCAGTAAATAGTCAAT 960
DB 1001 aaaaagctcgaacgtgcctcggagagcttaacctattcagagaataatgct 1060
OY 961 GGTACAGCTCTCAAAAGGTGAGCCATACCTATCGAAGATAGGGGAATTGAGTT 1020
DB 1061 ggtacagctctcctaaaggtgggccaatgactacgaagaatggtggaaattg 1120
OY 1021 GCCGATAGTGTGATTTGCTTTTAGGGAATACAGTCACTTACTACTCTGGAG 1080
DB 1121 gccgatagtgatgataatgtcttctttaggaataacagtaacttactacc 1180
OY 1081 AATAGAGTACTATGACTTAGAAGCAGTCGAAGATGACGCTTTCGCTTCTGCT 1140
DB 1181 aatagaagtaatgacttagaagcaggtgcgaagatgacagcttgcctgcgt 1240
OY 1141 GGTAGGCCATCTACTTATGATTCCTAATCCATTAACAGATCTTCCACAAAGTT 1200
DB 1241 ggtagagccattacttctatgaatcccaactacaagatccacaacagttaca 1300
OY 1201 GTCCTTAAAGTTAATGAGACTCCGGCAGATTCTGCATCTACATATACAGGAA 1260
DB 1301 gtcttaaaagttaatgaactcccgccagattctgcactacaatatacagga 1360
OY 1261 TTCACAGAGAAAAGTTATTCAGACAGACAGCCGAGATTCTTAAAAATCTTA 1320
DB 1361 ttcaagaagaaagttaatcagagcagaagccgcagaattcctaanaacttct 1420
OY 1321 CTACTACGCTGTACTCTTTCAGAGAGTACTCTATCTTAAAAATGAGTGCTG 1380
DB 1421 ctactacagcctgttaactcttccagaggtactcttacttaaacatgagtg 1480
OY 1381 CAGACTCAGGCTTCTCAACAGCAGCAGATTCTGCTCTGAAATGAGCTGAGAA 1440
DB 1481 cagactcagcattcactcaacacagcagattctcgtccgaatgagcgtgaac 1540
OY 1441 CTAGAACCTGCTGATTAAGCAGACATAAATTTGCTATTAACATCACTTATAG 1500
DB 1541 ctagaacctgctgatactagacccaataaacttgcatttaacatcagttcta 1600
OY 1501 GGTGCAAGAGGCAAAATATGAAACCAAACTCTGTCGAAAAATCTGATTAAT 1560
DB 1601 ggtgcaagagcgaataaataagaaacaaagctacgtaaaaaacttgaactc 1660
OY 1561 AACACTATTATTTGAGACCCAGCGGACGTTTATGAAAAATCATAGTTTAA 1620
DB 1661 aacactattatttgaacctcagcagcagcgttcttaagaaatcatagttta 1720
OY 1621 CAGTCTACGACATCTTAAGCTCAAGCTTCTGAGCTGTACAGACCGCAGTACT 1680
DB 1721 cagtcctcagacatcttagagctcacaagctcttgaactgttaacaagcac 1780
OY 1681 CCAGATCTATTAATGAGGAGAAATTCATTACGGCTTCAGGGAAGTTGGGCCA 1740
DB 1781 ccagatccttaataatggatgaaatccattcaagctacaggaacttggggcca 1840
OY 1741 GTTTGGGGAGACGGGCTTCTACGACTCAACCTTCAACTGGACTGAACTGAT 1800
DB 1841 gtttggggagacagggccttccagcgtcgaacttcaactcgaactgaata 1900
OY 1801 CCTAATCCGAGCGTATCGGCTCTTACTCCCTAATTAATGATGATTAATGAT 1860
DB 1901 cctaattcccgagcgtatcgcgtctttagtccctaataagcttaagatgata 1960
OY 1861 ATTACTCTCTCCATTAATCTATGAGACTGCAAGCAAGGTTTCAGGAGACCT 1920
DB 1961 attactctctccattacttatttgaagactgcaacgaaggcttgcagggacc 2020
OY 1921 TTTTGGTGGCTGATTTATCTTCACTTCTTCAATAGATAGTAAACAGCAGCG 1980
DB 2021 ttttgggtgctgatacttcaactcttccataagatagtaacaacaacagcgg 2080
OY 1981 TTTTCGCAATTTGAGTGGGTTATGTCATAGAGGAAACCTACATATCTTGTCA 2040
DB 2081 tttcgccaattgagcgggtatgtcatagaggaacactacatacttgcagata 2140
OY 2041 ATTCTTACTGCTGCAATTTTGCAAGCTCTTGGAAAGATAGACTATTGTTGCT 2100
DB 2141 attcttactgctgaatlltgcagctcttggaaagatagactacttgttagta 2200
OY 2101 AATCAGTACAGTCTACGAGAGAACTCTATTAACGCAACAGAACTATATCT 2160
DB 2201 aatcaagttacagcttaaggggaactctctatatacgaacaagaacctata 2260
OY 2161 CTTCCTGCAAACTACGCGCTTGTTCGTTGCTTATGTTCTCTACAGAGATTCT 2220
```



```

QY 1681 CCAGATCCTTAATGAGTGGAGAAATTCATTACGGCTATACAGGAACTGGGGCCCAATT 1740
Db 1681 ccagatccttaataatgagtgagaatctcattacggctatacagggaacttgggcccatt 1740
QY 1741 GTTTGGGGGACAGGGGCTTCTACGACGTCGCAACTTCAACTGAGCTAAATCGCTATATT 1800
Db 1741 gtttgggggacaggggcttctacgacgtcgcgcaacttcaactgagctaaatcgctataat 1800
QY 1801 CCTAATCCCGGCGTATCCGGCTCTTTAGTCCCTAATAGCTTATGGAATGCAATTTATAGAT 1860
Db 1801 cctaatacccgagcgtaacgctctttagccctaataagcttaagaaagcattatagat 1860
QY 1861 ATTAGCTCTCCATTAATCTTATGAGACTGCAAAACGAGGTTGCAAGGAGACCGTCTCT 1920
Db 1861 attagctctccattatctatagagactgcaaaagaggttgcaggagaccggtct 1920
QY 1921 TTTTGGTGTGCTGATTAATCTTAATCTTCCATAGAGTATGACAAAAACGACGCGGG 1980
Db 1921 ttttgggtgtgctgattatcttaactcttccataagtagtaaaaaacgacgcggg 1980
QY 1981 TTTTGGCATTTGAGTGGCGGTTATGTCATAGAGGAAACCTACATCTGTTTCAGATAG 2040
Db 1981 ttttggcatTTGAGTGGCGGTTATGTCATAGAGGAAACCTACATCTGTTTCAGATAG 2040
QY 2041 ATTCTAGTCTGCTGATTTGTCACCTCTTTGAGAGATAGAGACTTCTTGTAGCTAAG 2100
Db 2041 attcttagtctgctgatttctgcaagctcttggagaagatagagactcttctgataag 2100
QY 2101 AATCAGGTACAGTCTACGAGAGAACTCTATTACACAGCAACGAAACCTATATCTCT 2160
Db 2101 aatcaggtacagctctacgaggaactctctattacagcacaagaaacctatactct 2160
QY 2161 CTTCCTTGCAAACTACGGCCCTTGTGTCGTCATATGTCCTACAGAGATTCGCTCTCTC 2220
Db 2161 ctctcttgcaaaactacgagccctgtctgtctctatgtctctacagagatctccgtctctc 2220
QY 2221 TTTTTCAGAAACCTTACCTACACCCATACGAGATTAACGATCTGAAACCAATATACAA 2280
Db 2221 tttttcagaaaccttactacacccatacagataagatctgaaacaagtatacaaca 2280
QY 2281 TATCTTACTGTTAAAGAGAGCTGGGGAAATGATAGTTTCGCTTAAATTCGTTGAGAA 2340
Db 2281 tatcttactgttaaaggaagctggggaatgatatgtctgctttagaatctcgtggagaa 2340
QY 2341 GCTCCGATTTGCTATAGTAAAGTGCCTTTTGAAGAGTACAGCCCTCATCAAAATTTG 2400
Db 2341 gctccgatttgcctatagaaagtgctctatcttgagcagtaacatgcccctcaatgaatcg 2400
QY 2401 CAGTTTGTCTATGACATACAGAGAGGTTTAAAGACAGGAGACAGAGCTCGTGAATTT 2460
Db 2401 cagtttgtctatgacatcacgaaaggttttaaagaaacaggaagaaagctcgtgaattc 2460
QY 2461 GGAAGTACCGCTCTTGGATCTTGCTTACCTATTCGGGATCGAATTTGATTAAGAAATCA 2520
Db 2461 ggaagtacccgctcttgatcttgcttactcctacggaatcgaattgataaagaatca 2520
QY 2521 GACGCGCAAGTGAAGCTCAATCTAACCTTGTTATCTGTTATCTGTTGCTGAGT 2580
Db 2521 gacgcgcaagtgaagctaaactcaactccttggtatactgtagatcttgcctgtagt 2580
QY 2581 AACCCGACGCTACAGACACTGCGAATTAAGCGTGAATTTCTTGGAAACCTTCGTCAGC 2640
Db 2581 aaccgcgactgaacaaacactcgcgaattagcggtgattcttggaaaacctcgtgacg 2640
QY 2641 AATTTGGCAAGACAGCTTTATGCTTCGTCGAGGAGAACATTTTCTTACTCAAAAT 2700
Db 2641 aatttggcaagacagctttatgcttcctcgtcagggagacatttcttctaatacaat 2700
QY 2701 TTTGAACGCTTTGACCAATTTCTTTGAATTTGGTGGGTCATCGCAATTAACAAGTA 2760
Db 2701 ttctgaacgctttgaccaatttcttcttgaatttgcgtggtcattctcgcaattacaatgta 2760
QY 2761 GACTTAGAGACAAATATCAATTTCTAA 2787

```

```

Db 2761 gacttagagcaaaataccaattcttaa 2787

RESULT 2
ID AAA27021 standard; DNA; 3000 BP.
AC AAA27021;
XX
XX
XX 11-AUG-2000 (first entry)
XX
XX Chlamydia pneumoniae 98kD putative outer membrane protein gene.
XX Chlamydia; antigen; vaccine; infection; outer membrane protein; ds.
XX
XX Chlamydia pneumoniae.
XX
XX Key Location/Qualifiers
XX Key 101..2887
XX CDS /*tag= a
XX /product= "98kDa putative outer membrane protein"
XX
XX W0200026237-A2.
XX
XX 11-MAY-2000.
XX
XX 29-OCT-1999; 99WO-GB03579.
XX
XX 29-OCT-1998; 98US-0106070.
XX 01-MAR-1999; 99US-0122066.
XX 27-OCT-1999; 99US-0428122.
XX
XX (CONN-) CONNANUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Dunn PL;
XX WPI: 2000-365569/31.
XX DR P-PSDB; AAY94327.
XX
XX Novel Chlamydia 98 kDa putative outer membrane protein antigen, used
XX PT for vaccination and protection against Chlamydia infection
XX
XX Claim 1; Fig 1; 93pp; English.
XX
XX The present sequence is the 98kDa putative outer membrane protein gene
XX from Chlamydia pneumoniae. The genomic sequence was amplified using two
XX PCR primers. The 5' primer contains a NciI restriction site, a ribosome
XX binding site, an initiation codon and a sequence close to the 5' end of
XX the 98kDa putative outer membrane protein coding sequence. The 3' primer
XX contains the sequence encoding the C-terminal sequence of the putative
XX outer membrane protein and a BamHI restriction site. The stop codon was
XX excluded and an additional nucleotide was inserted to obtain an in-frame
XX C-terminal fusion with the Hisidine tag. The PCR product was cloned
XX into a eukaryotic expression vector (pCA-Myc-His) by restricting both
XX the vector and the PCR product with NotI and BamHI and performing a
XX ligation reaction. This expression vector was injected intramuscularly
XX and intranasally into mice, which were subsequently inoculated with
XX Chlamydia pneumoniae. The chlamydial lung titers of the immunised mice
XX were lower than those of the controls. Thus the 98kDa putative outer
XX membrane protein can be used as a vaccine to provide protection against
XX Chlamydia infections, especially Chlamydia pneumoniae infections.
XX The polypeptide may also be administered orally to treat Chlamydia
XX infection. The present sequence may also be used in the
XX construction of attenuated Chlamydia strains that can over-express the
XX gene or express it in a non-toxic form.
XX
XX Sequence 3000 BP; 871 A; 617 C; 634 G; 878 T; 0 other;

```

```

Query Match 99.6%; Score 2775.8; DB 21; Length 3000;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2780; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 05:32:59 ; Search time 800.41 Seconds  
(without alignments)  
2186.332 Million cell updates/sec

Title: US-09-446-677B-9  
Perfect score: 2787  
Sequence: 1 ATGCAAGCTCTCTTCCCA.....GAGCAAAATACCAATTTCTAA 2787

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N.Geneseq\_0601:\*

1:	/SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT:*
22:	/SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2787	100.0	2787	20	AAK06820
2	2775.8	99.6	3000	21	AAK27021
3	2760.8	99.1	1230025	20	AAK91990
4	536	19.2	273254	21	AAK81914
5	535	19.2	2815	20	AAK06817
6	528.6	19.0	2757	20	AAK06821
7	526.8	18.9	2957	21	AAK61509
8	518.6	18.6	2950	21	AAK30851
9	517.4	18.6	2950	21	AAK30849
10	517	18.6	2787	20	AAK06822
c 11	517	18.6	273254	21	AAK81914

12	515.6	18.5	2784	21	AAA30852	Chlamydia antigen
13	514	18.4	2793	20	AAK06823	Chlamydia pneumoni
14	512.4	18.4	3000	21	AAK30853	Chlamydia antigen
15	510.8	18.3	3000	20	AAK06828	Chlamydia pneumoni
16	497.2	17.8	3050	21	AAA48838	Chlamydia antigen
17	491.8	17.6	2790	21	AAA30854	Chlamydia antigen
18	474.2	17.0	2784	21	AAA30850	Chlamydia antigen
19	470.6	16.9	2808	21	AAA48839	Chlamydia pneumoni
20	465.8	16.7	2950	21	AAA30847	Chlamydia antigen
21	465.8	16.7	3200	20	AAK06816	Chlamydia antigen
22	435.8	15.6	2784	21	AAA30848	Chlamydia antigen
23	389.6	14.0	3050	20	AAK61508	DNA encoding the C
24	380	13.6	2838	20	AAK06827	Chlamydia pneumoni
25	343.4	12.3	2526	21	AAA28691	C. pneumoniae CPN1
26	343.4	12.3	2750	21	AAA28690	C. pneumoniae CPN1
27	338.6	12.1	2526	20	AAK06819	Chlamydia pneumoni
28	270	9.7	3052	20	AAK06818	Chlamydia pneumoni
29	268.6	9.6	2950	21	AAA50035	DNA encoding Chlam
30	244.8	8.8	3150	21	AAA27342	Chlamydia POM91B
31	207.6	7.4	1830	20	AAK06830	Chlamydia pneumoni
32	197.6	7.1	840	20	AAK06824	Chlamydia pneumoni
33	139	5.0	3354	20	AAK60556	Seq ID No. 23 of W
34	132.2	4.7	3021	21	AAK64758	C. trachomatis pmp
35	132.2	4.7	3324	21	AAK60557	C. trachomatis pmp
36	131.4	4.7	2949	21	AAK64752	C. trachomatis pmp
37	130.6	4.7	1444	20	AAK60547	C. trachomatis LGV
38	130.6	4.7	4435	20	AAK60539	Chlamydia pneumoni
39	104	3.7	1200	20	AAK06829	Chlamydia pneumoni
c 40	88.4	3.2	936	22	AAK58252	Oligonucleotide D1
c 41	88.4	3.2	936	22	AAK58254	Oligonucleotide D1
c 42	88.4	3.2	936	22	AAK58257	Oligonucleotide D1
c 43	88.4	3.2	936	22	AAK58259	Oligonucleotide D1
c 44	88.4	3.2	936	22	AAK58262	Oligonucleotide D2
c 45	88.4	3.2	938	22	AAK58255	Oligonucleotide D1

## ALIGNMENTS

RESULT	1
AAK06820	AAK06820 standard; DNA: 2787 BP.
ID	AAK06820:
XX	26-APR-1999 (first entry)
AC	Chlamydia pneumoniae surface exposed protein Omp8 DNA.
XX	Omp8: outer membrane protein 8; surface exposed protein; antigen;
XX	Infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
KW	Chlamydia pneumoniae.
OS	Chlamydia pneumoniae.
XX	MO9858953-A2.
PN	30-DEC-1998.
XX	19-JUN-1998; 98MO-DK00266.
PF	23-JUN-1997; 97DK-0000744.
PR	(BIRK/) BIRKELUND S.
XX	(CHRI/) CHRISTIANSEN G.
PA	BIRKELUND S, Christiansen G, Knudsen K, Madsen A;
PA	Mygdal P;
PI	WPI, 1999-105610/09.
XX	P-PSDB; AAK88421.
DR	Species-specific test for identifying mammals infected with
XX	Chlamydia pneumoniae - comprises detecting antibodies specific for
PT	Chlamydia pneumoniae

**This Page Blank (uspto)**

***This Page Blank (uspto)***

**BEST AVAILABLE COPY**





```

?       TELEX:
?       INFORMATION FOR SEQ ID NO: 9
?
?       SEQUENCE CHARACTERISTICS:
?           LENGTH: 5340
?           TYPE: nucleic acid
?           STRANDEDNESS: double
?           TOPOLOGY: linear
?       MOLECULE TYPE: genomic DNA
US-08-492-459-9

```

Query Match	1.2%	Score 33.6	DB 3	Length 5340
Best Local Similarity	45.7%	Pred. No. 6	5	
Matches 117	Conservative 0	Mismatches 139	Indels 0	Gaps 0

Qy	435	TACGGGATCGTTAAACATTGACAAAAATGCTGCTTCACAAAATCTTTTCAAC	494
Db	3512	TACAGATAGCTTNGATTAATAGAGTAAACGAAAGTTTACGCTTGATGACATCTCAATTTGC	3571
Qy	495	GGATATATGCGGCTGCTATACCGCAAAAACCTTTTCATTACAGGAGTACATGTCAAC	554
Db	3572	TAAATATGTAAGGTGTGGGTATGATGTGTGTGTTACTTGCCGTTGGTGTATGCAAT	3631
Qy	555	TCTGTTTTCGAATAATACCCCTCAAGAAAGGCGAGCATTCACAGCTTCGATGCCCT	614
Db	3632	TCCGTTTCTTTTGGTCATCTTTTCTGTGATTCGTATCATTTATCAGAGTTCGTGATGAGA	3691
Qy	615	TACCAATTACTGGAACCAAGGGGAAGTCTCTTTTTCGACAATACTCTTCGGATTGGG	674
Db	3692	AAATTAAGAAGCTTGAAGCTGTGCACGCGCTTTTGTTCACATTAATTTAAATGAAGTTT	3751
Qy	675	AGCAGCAATTTTACA	690
Db	3752	GGGTGGGATGGATACA	3767

RESULT 11  
 US-08-423-752-9  
 ; Sequence 9, Application US/08423752  
 ; Patent No. 6022949  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takashi OKADO et al.  
 ; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING  
 ; TITLE OF INVENTION: ADREOBASIDIN SENSITIVITY  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Wenderoth, Lind & Ponack  
 ; STREET: 805 Fifteenth Street, N.W., #700  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: U.S.A.  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: Wordperfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/423,752  
 ; FILING DATE: April 18, 1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/243,403  
 ; FILING DATE: May 16, 1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Warren M. Cheek, Jr.  
 ; REGISTRATION NUMBER: 33,367  
 ; REFERENCE/DOCKET NUMBER:  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-371-8850  
 ; TELEFAX:  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 9:

```

: SEQUENCE CHARACTERISTICS:
:     LENGTH: 5340
:     TYPE: nucleic acid
:     STRANDEDNESS: double
:     TOPOLOGY: linear
:     MOLECULE TYPE: genomic DNA
:
US-08-423-752-9

```

Query Match	1.2%	Score 33.6;	DB 3;	Length 5340;
Best Local Similarity	45.7%	Pred. No. 6.6;		
Matches 117; Conservative	0;	Mismatches 139;	Indels 0;	Gaps 0;

QY	435	TACGGGAGCTTAAAGTTGACAAATAATGTCAGTTTGCCTTCAGCAAAAACCTTTTCAAC	494
Db	3512	TACAGATAGCTTAGATATATGATAGTTAAACCCGAAAGTTACGCTTGATGACATCTCAATTTGC	3572
QY	495	GGATAAATGGGGGTGCATACACGCCAAAACCTTTTATTAACAGGACATACAATGTACG	554
Db	3572	TAAATATGATAGTGTGTGGCTCATGTGATATGTTAACTTGCCGTGGTTGGATATGCAAAT	3633
QY	555	TCGTGTTTTGAAAATATACCTCCCAAGAAAGGCGAGCCATTCAGACTTCGCATGCCCT	614
Db	3632	TCCGTTCTTTTGGTCATCTTTTGTCTGATGTATCATTAATACAGAGTTCGGATAGAGA	3692
QY	615	TACCATTTACTGGAACCAAGGGAAGTCTTTTTCGCAATACCTCTTCGCGATTTGCG	674
Db	3692	AATTTAAAGACTTGAAGCCTGTGCACGCGCTTTTGTTCACATTAATTTAAATGAAGTTT	3751
QY	675	AGCTCAATTTTACA	690
Db	3752	GGGTGGGATGATACA	3767

RESULT 12  
 US-08-945-994-2  
 : Sequence 2, Application US/08945994  
 : Patent No. 6043051  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: Takashi OKADO et al.  
 : TITLE OF INVENTION: PROMOTER  
 : NUMBER OF SEQUENCES: 8  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
 : STREET: 2033 K Street, N.W., #800  
 : CITY: Washington  
 : STATE: D.C.  
 : COUNTRY: U.S.A.  
 : ZIP: 20006  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
 : COMPUTER: IBM Compatible  
 : OPERATING SYSTEM: MS-DOS  
 : SOFTWARE: Wordperfect 5.1  
 :  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/945,994  
 : FILING DATE: No. 6043051ember 6, 1997  
 : CLASSIFICATION: 435  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER:  
 :  
 : FILING DATE:  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Warren M. Cheek, Jr.  
 : REGISTRATION NUMBER: 33,367  
 : REFERENCE/DOCKET NUMBER:  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 202-721-8200  
 : TELEFAX:  
 :  
 : TELEX:  
 : INFORMATION FOR SEQ. ID NO.: 2:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 5340  
 : TYPE: nucleic acid



TITLE OF INVENTION: MUTATIONS IN THE APC GENE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER & WITCOFF, LTD.  
STREET: 1001 G STREET, N.W.  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: US  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/370,235A  
FILING DATE: 01-JAN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107.48688  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 508 9100  
TELEFAX: 202 508 9299  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9606 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
MOLECULE TYPE: cdna  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
POSITION IN GENOME:  
MAP POSITION: 5q21  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 34..8562  
US-08-370-235A-1

Query Match 1.3%; Score 35.6; DB 2; Length 9606;  
Best Local Similarity 57.0%; Pred. No. 2.3;  
Matches 65; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 1837 AGCTTATGATGATGATTTATGATATGCTCTCCATTTATCTTATGAGAGTGAAC 1896  
Db 793 AGCGGTTTCACTGTTGTTGAGATGACCTCTCTGTTGTTGTTGAGACTGTAAA 734  
Qy 1897 GAAGGTTGACGAGACCGTCTTTTGTGTCCTGATTAATCTTACTCTTC 1950  
Db 733 GCTGCTGATACGAAGATGATGCTTTGATTTGCTGAATTCGTGATCTTTC 680

RESULT 9  
US-09-157-210-10  
Sequence 10, Application US/09157210B  
Patent No. 6204003  
GENERAL INFORMATION:  
APPLICANT: Steele, J. Kevin  
APPLICANT: Telford, David L.  
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS OF FELINE  
TITLE OF INVENTION: INFECTIOUS ANEMIA  
FILE REFERENCE: SYNBO.100A  
CURRENT APPLICATION NUMBER: US/09/157,210B  
CURRENT FILING DATE: 1998-09-18  
EARLIER APPLICATION NUMBER: 60/059,551  
EARLIER FILING DATE: 1997-09-19  
NUMBER OF SEQ ID NOS: 15

SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 10  
LENGTH: 977  
TYPE: DNA  
ORGANISM: Haemobartonella felis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3)...(80)  
NAME/KEY: CDS  
LOCATION: (81)...(365)  
NAME/KEY: CDS  
LOCATION: (366)...(419)  
NAME/KEY: CDS  
LOCATION: (420)...(944)  
NAME/KEY: CDS  
LOCATION: (945)...(977)  
NAME/KEY: misc\_feature  
LOCATION: 186, 816, 860, 885, 952, 959  
OTHER INFORMATION: n = A, T, C or G  
US-09-157-210-10

Query Match 1.2%; Score 33.6; DB 4; Length 977;  
Best Local Similarity 51.0%; Pred. No. 2.4;  
Matches 78; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 2329 TTCGGTGAAGAGCTCCCATTTGCTTATGATGAAGTCTCTATTTGACGATACATGCC 2388  
Db 57 ttccagaccatcccatatgactactaaagaagcctaactcagcagatgaaag 116  
Qy 2389 TTCAATGAATTTGAGTTGTATGACATGACATGAGAGGTTTAAAGACAGGAACAGAA 2448  
Db 117 gtttgaatcatctattatagatgacacacacacacacatattgaagatgattagaa 176  
Qy 2449 GCTCGTAATTTGGAAGTACGCCGTTCTGTGAAT 2481  
Db 177 gctaagtaantagaataaagactcttgaat 209

RESULT 10  
US-08-492-459-9  
Sequence 9, Application US/08492459  
Patent No. 6015689  
GENERAL INFORMATION:  
APPLICANT: Takashi OKADO et al.  
TITLE OF INVENTION: REGULATION OF AUREOBASIDIN SENSITIVITY IN FUNGUS  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/492,459  
FILING DATE: June 20, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/243,403  
FILING DATE: May 16, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX:

APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YTSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Allegretti, LTD  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/289,548A  
FILING DATE: 12-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107,46943  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9606 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: DP2.5(APC)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 34..8562  
US-08-289-548A-1

Query Match 1.3%; Score 35.6; DB 1; Length 9606;  
Best Local Similarity 57.0%; Pred. No. 2.3;  
Matches 65; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 1837 ACCTATGGAATGCAATTATAGATATAGCTCTCCATTAATCTTATGAGAGCTGCAAC 1896  
DB 793 ACCCGGTTTCATGCTTGTCTGAGATGACCTCTGCTCTGCTGCTGGAGCTGTAATA 734  
QY 1897 GAAGGTTCCAGAGAGACCGCTTTTGGTGGCTGATTAATCTTACTTCTTC 1950  
DB 733 GCTGTCGATACGAAGTATGTCCTTTCGATTGCTGAATTCGGCTATTCCTTC 680

RESULT 7  
US-08-452-654-1/C  
; Sequence 1, Application US/08452654  
; Patent No. 561454  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY  
; APPLICANT: GRODEN, JOANNA  
; APPLICANT: HEDGE, PHILIP J.  
; APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YTSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,654  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107,035574  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9606 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: DP2.5(APC)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 34..8562  
US-08-452-654-1

Query Match 1.3%; Score 35.6; DB 1; Length 9606;  
Best Local Similarity 57.0%; Pred. No. 2.3;  
Matches 65; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 1837 ACCTATGGAATGCAATTATAGATATAGCTCTCCATTAATCTTATGAGAGCTGCAAC 1896  
DB 793 ACCCGGTTTCATGCTTGTCTGAGATGACCTCTGCTCTGCTGCTGGAGCTGTAATA 734  
QY 1897 GAAGGTTCCAGAGAGACCGCTTTTGGTGGCTGATTAATCTTACTTCTTC 1950  
DB 733 GCTGTCGATACGAAGTATGTCCTTTCGATTGCTGAATTCGGCTATTCCTTC 680

RESULT 8  
US-08-370-235A-1/C  
; Sequence 1, Application US/08370235A  
; Patent No. 5910418  
; GENERAL INFORMATION:  
; APPLICANT: VOGELSTEIN, BERT  
; APPLICANT: KINZLER, KENNETH W.  
; APPLICANT: HILL, DAVID E.  
; APPLICANT: JOHNSON, KAREN A.  
; TITLE OF INVENTION: ANTIBODIES AND ASSAYS FOR DERMINING

APPLICANT: JOSLYN, GEOFF  
 APPLICANT: KENNETH  
 APPLICANT: MARKHAM, ALEXANDER F.  
 APPLICANT: NAKAMURA, YUSUKE  
 APPLICANT: THILVERIS, ANDREW  
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATION  
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN  
 NUMBER OF SEQUENCES: 102  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Banner & Witcoff, Ltd.  
 STREET: 1001 G Street, NW  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20001-4598  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/450,582  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/452,655  
 FILING DATE: 25-MAY-1995  
 APPLICATION NUMBER: US 08/289,548  
 FILING DATE: 12-AUG-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/741,940  
 FILING DATE: 08-AUG-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kagan, Sarah A.  
 REGISTRATION NUMBER: 32,141  
 REPRESENTATION/DOCKET NUMBER: 1107.49964  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-508-9100  
 TELEFAX: 202-508-9299  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 8532 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: DP2.5(Apc)  
 US-08-450-582-1

	Query Match	1.3%	Score 35.6;	DB 3;	Length 8532;
	Best Local Similarity	57.0%;	Pred. No. 2.2;		
	Matches 65;	Conservative	0;	Mismatches 49;	Indels 0;
				Gaps	0;
QY	1837 AGCTATGGAATGCATTATAGATATTTACCTCTCCATTACTTATGTGAGAGACTGCACAAAC				1896
Db	760 AGCCGGTTTCATGCTTGCTTCGAGATGACACCTCTCGCTTCCTTGCGTGGAGCACTGTAAAA				701
QY	1897 GAAGGCTGGCAGGGGAGACCGTCCTTTTGGTGTGCGTGATATCTCAACTTCTTC				1950
Db	700 GCTGCGTATACGAAGATATGCTCTTTTGATTGGCATTTGGCTATATCTTC				647

RESULT 5  
US-07-741-940-1/c  
: Sequence 1, Application US/07741940  
: Patent No. 5352775  
: GENERAL INFORMATION:  
: APPLICANT: ALBERTSEN, HANS  
: APPLICANT: ANAND, RAKESH

```

APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSTLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YOSUKE
APPLICANT: THLIVERTS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/741,940
FILING DATE: 19920109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9606 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: DP2.5(APC)
FEATURE:
NAME/KEY: CDS
LOCATION: 34..8562
IS-07-741-940-1

```

	Query March	Best Local Similarity	1.3%, 57.0% Pred. No. 2.3	Score 35.6	DB 1	Length 9606
	Matches	Conservative	0	Mismatches	49	Indels
					0	Gaps
QY	1837	AGCTTATGGAATGCATTATAGATATTTAGCTCTCCACATTA	CTGAGAGCTGCAAC	1896		
Db	793	AGCCGGTTTCATGCTGTTCGTGAGATGACCTCTGCGCTCTCTTCTC	CTGGGGAGCTGTAATA	734		
QY	1897	GAAGGGTTGCAGGGAGACCGTGCCTTTTGGTGTGCGGATTA	CTTAACCTCTTC	1950		
Db	733	GCTGTCGTATACGAAGTATGTCCTTTTGATTTGCTGATATTC	GCTATATTC	680		

RESULT 6  
US-08-289-548A-1/c  
Sequence 1, Application US/08289548A  
Patent No. 5668212  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: NAAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA

```
OY 624 TGGAAACGAGGAGTCTTTTCTGACATACCTCTCGGATTCGAGCTGCAT 683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 TGGGACCATGGGAGTACACGATGTGATGTCCTCCAGTTCGGGATTTGCT 458
OY 684 TTTTACAGAGGCTCGGTACTTTCTAATATGCTAAAGTTCTTTATTGACATA 743
    - - - - - ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 459 CAGAAATATGAGCTCGATCTGATCTTCAATACGAAAGCTGCTACTCCATCAT 518
OY 744 GGTACAGAGGAGCTCTCAACAGGGGATATGTCAGAGTCTATCTGTGCTTA 803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 519 TGACAAACTCCGTACTACTAGAAATGAGAGATTGCTACTCTTAAAGGAAAA 578
OY 804 TAAACTAGTACAGATCTA 823
    ||| ||| ||| |||
Db 579 AGAACCCGGCTCAATGTTA 598

RESULT 2
US-09-107-858-11
; Sequence 11, Application US/09107858
; Patent No. 6162900
; GENERAL INFORMATION:
; APPLICANT: Gueriot, Mary Lou et al.
; TITLE OF INVENTION: METAL-REGULATED TRANSPORTERS AND USES THEREFOR
; FILE REFERENCE: DCI-099CPV
; CURRENT APPLICATION NUMBER: US/09/107,858
; EARLIER FILING DATE: 1998-06-30
; EARLIER APPLICATION NUMBER: 08/758,621
; EARLIER FILING DATE: 1996-11-27
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1269
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1266)
; US-09-107-858-11

Query Match 1.3%; Score 36.8; DB 4; Length 1269;
Best Local Similarity 49.0%; Pred. No. 0.3; Mismatches 102; Indels 0; Gaps 0;
Matches 98; Conservative 0;
```

```
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, AYUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,655B
FILING DATE: 25-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: DP2.5(APC)
US-08-452-655B-1

Query Match 1.3%; Score 35.6; DB 1; Length 8532;
Best Local Similarity 57.0%; Pred. No. 2.2; Mismatches 49; Indels 0; Gaps 0;
Matches 65; Conservative 0;
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 03:46:39 ; Search time 268.88 Seconds  
(Without alignments)  
1962.251 Million cell updates/sec

Title: US-09-446-677b-9  
Perfect score: 2787  
Sequence: 1 ATGAGGTCTTCTTCCCA.....GAGCAAAATACCAATCTTAA 2787

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_patents\_MA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCYUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36.8	1.3	1269	2 US-08-758-621-11	Sequence 11, Appl
2	36.8	1.3	1269	4 US-09-107-858-11	Sequence 11, Appl
3	35.6	1.3	8532	1 US-08-452-655B-1	Sequence 1, Appl
4	35.6	1.3	8532	3 US-08-450-582-1	Sequence 1, Appl
5	35.6	1.3	9606	1 US-07-741-940-1	Sequence 1, Appl
6	35.6	1.3	9606	1 US-08-289-548A-1	Sequence 1, Appl
7	35.6	1.3	9606	1 US-08-452-654-1	Sequence 1, Appl
8	35.6	1.3	9606	2 US-08-370-235A-1	Sequence 1, Appl
9	33.6	1.2	977	4 US-09-157-210-10	Sequence 9, Appl
10	33.6	1.2	5340	3 US-08-492-459-9	Sequence 9, Appl
11	33.6	1.2	5340	3 US-08-423-752-9	Sequence 9, Appl
12	33.6	1.2	5340	3 US-08-945-994-2	Sequence 2, Appl
13	33.6	1.2	7218	1 US-08-716-873-23	Sequence 23, Appl
14	33.2	1.2	7218	1 US-08-232-463-14	Sequence 14, Appl
15	32.8	1.2	3330	1 US-08-149-103-1	Sequence 1, Appl
16	32.8	1.2	3330	1 US-08-451-883-1	Sequence 1, Appl
17	32.6	1.2	1357	2 US-08-546-117-8	Sequence 8, Appl
18	32.2	1.2	246240	2 US-08-724-394A-20	Sequence 20, Appl
19	32.2	1.2	246240	2 US-08-724-394A-21	Sequence 21, Appl
20	32.2	1.2	246240	2 US-08-724-394A-22	Sequence 22, Appl
21	31.8	1.1	448	1 US-08-702-080-7	Sequence 7, Appl
22	31.8	1.1	448	2 US-08-858-830-7	Sequence 7, Appl
23	31.8	1.1	448	2 US-08-858-834-7	Sequence 7, Appl
24	31.6	1.1	7295	2 US-08-487-826B-15	Sequence 15, Appl
25	31.4	1.1	1696	3 US-09-028-366-1	Sequence 1, Appl
26	31.2	1.1	3656	1 US-08-393-734-1	Sequence 1, Appl
27	31.2	1.1	3656	4 US-08-894-489-1	Sequence 1, Appl

c 28	31.2	1.1	5049	1 US-08-336-345-1	Sequence 1, Appl
c 29	31.2	1.1	5049	2 US-08-336-345-2	Sequence 2, Appl
c 30	31.2	1.1	5049	1 US-08-647-655-1	Sequence 1, Appl
c 31	31.2	1.1	5049	2 US-08-647-655-2	Sequence 2, Appl
c 32	31.2	1.1	9592	1 US-08-393-734-3	Sequence 3, Appl
c 33	31.2	1.1	9592	4 US-08-894-489-3	Sequence 3, Appl
c 34	31.2	1.1	1560	2 US-08-794-795-5	Sequence 5, Appl
c 35	31.2	1.1	1560	4 US-09-249-200-5	Sequence 5, Appl
c 36	31.2	1.1	1703	2 US-08-794-795-1	Sequence 1, Appl
c 37	31.2	1.1	1703	4 US-09-249-200-1	Sequence 1, Appl
c 38	30.8	1.1	1985	1 US-07-792-865D-2	Sequence 2, Appl
c 39	30.8	1.1	2694	3 US-08-975-703-5	Sequence 5, Appl
c 40	30.8	1.1	2694	4 US-09-515-884-5	Sequence 5, Appl
c 41	30.6	1.1	1473	4 US-08-907-740-6	Sequence 6, Appl
c 42	30.6	1.1	6763	2 US-08-756-506-23	Sequence 23, Appl
c 43	30.6	1.1	6924	1 US-08-015-973-2	Sequence 2, Appl
c 44	30.6	1.1	6924	2 US-08-448-164-2	Sequence 2, Appl
c 45	30.6	1.1	10807	1 US-08-206-176-7	Sequence 7, Appl

#### ALIGNMENTS

RESULT 1  
US-08-758-621-11  
Sequence 11, Application US/08758621  
Patent No. 5846821  
GENERAL INFORMATION:  
APPLICANT: Guerintot, Mary Lou, and Elide, David J.  
TITLE OF INVENTION: Metal-Regulated Transporters and Uses Therefor  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/758, 621  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/018, 578  
FILING DATE: 29-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Silverl, Jean M.  
REGISTRATION NUMBER: 39,030  
REFERENCE/DOCKET NUMBER: DCT-099CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1269 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1267  
US-08-758-621-11

Query Match 1.3%; Score 36.8; DB 2; Length 1269;  
Best Local Similarity 49.0%; Pred. No. 0.3;  
Matches 98; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Search completed: October 2, 2001, 03:06:59  
Job time: 51677 sec





Query Match	1.48;	Score 39;	DB 247;	Length 929;
Best Local Similarity	51.68;	Pred. No. 8.3;		
Matches 113;	Conservative	0;	Mismatches 105;	Indels 1;
				Gaps 1;
QY 1338	TCCTTCAGGAGCTACTCTATCTTTAAACATGAGGTGACTCTGCAGACTCAGGACTTCAC	1397		
Db 535	TTTTTCAGGTGTATTTTTTTCTAAATATCGTTGATTTACTTACATCATATGATGTTC	594		
OY 1398	TCAACGAGCAGATCTCGCTCGCAATGAGAGCTAGGACTACTCAGAACCGTCGATTC	1457		
Db 595	TGAAAATGCTATATCTCTCTCTTTAAATGCAAAATTAATTGATCAATTAATTTTAAAGTC	654		
OY 1458	TAGCACCATAAACAATTGGTCATTAAACATCAGTCTCTATAGCGGTGCAGAAAGAGCAAA	1517		
Db 655	T-CCTTTCTCAACACACTTTTGAATTACATGATTTCCATTTTGGTCTTGAACATGCAAT	713		
OY 1518	AATGAAACCAACCTACGTCAAAAAAATCTGACTTATC	1556		
Db 714	AATATTAATCTCTATTTTCATTAATAAATGATCTTTTATC	752		

RESULT	11
A1830489/c	
LOCUS	
DEFINITION	A1830489 372 bp mRNA
	w515106.x1 NC1_CGAR_Kd11 Homo sapiens CDNA clone IMAGE:384314 3', EST
ACCESSION	mRNA sequence.
VERSION	A1830489
KEYWORDS	A1830489.1 GI:5451160
SOURCE	EST.
	human.

ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (Bases 1 to 372)
REFERENCE AUTHORS TITLE	NCI-CCAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . National Cancer Institute, Cancer Genome Anatomy Project (CCAP), Tumor Gene Index
JOURNAL COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D.

Email: cgapds-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome  
Clone distribution: NCI-cGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LENI at: [www-bio.llnl.gov/bdrp/image/image.html](http://www-bio.llnl.gov/bdrp/image/image.html)  
Insert length: 393 Std Error: 0.00  
Seq primer: -400P from Glbco.

FEATURES	SOURCE
Location/Qualifiers	
1. 372	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone_image="IMAGE:2384314"	
/clone_11b="NCI CGAP_Kid11"	
/lab_host="DH10b"	
/note="Organ: kidney; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site: 1. Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Kid3 was prepared, and ss circles were made in vitro. Following HAT hybridization, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1342376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."	
131 a	
74 c	
55 g	
110 t	
2 others	

Query Match	1.4%;	Score 38.8;	DB 102;	Length 372;
Best Local Similarity	49.3%;	Pred. No. 6.8;		

	Matches	100;	Conservative	0;	Mismatches	103;	Indels	0;	Gaps	0;
OY	2172	ACTACGCGCTTGTCGTGTCTTANGTTCCTPACAGAGATTCCTGTTCCTTTTCAGGAAA	2231							
Db	288	ACAGAGGTTTTTTTAGTTTATTATGTGCACACAGAAATGTAACAATGTTTNCNTGATAAGAAAA	229							
OY	2232	CCTTAGTCACACCCTACGAGATACGATCTGAAAAACAAGTATPACACATATCCTACTGT	2291							
Db	228	CTCCACACTAACCTAAGAAGCATATTTAAATAGTTTAAAGTCTGTAATCTCCCCTTTTNG	169							
OY	2292	TAAAGGAAGCTGGGGAAATGATAGTTTCCGCTTTAGCAATTCGGTGGACAGCTCCGATTGG	2351							
Db	168	AAAAACAACCCCTGAGGGTTTTTTCGTGATGTGAAGGAGCGGTCTTTTGTGCCAGTNG	109							
OY	2352	CTTAGATCAAGAAGTCTATTTG	2374							
Db	108	TTTTCTTGTGTGGTGGTGGTGTG	86							

RESULT	12
LOCUS	C23720/c
DEFINITION	C23720 658 bp mRNA EST 16-OCT-1997
ACCESSION	C23720 Dictyostelium discoideum FC (H.Trushihara) Dictyostelium
VERSION	discoidaeum cDNA clone FC-AB21, mRNA sequence.
KEYWORDS	C23720 GI:2073232
SOURCE	EST.
ORGANISM	Dictyostelium discoideum.
	Dictyostelium discoideum
	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.

REFERENCE	1 (bases 1 to 658)
AUTHORS	Shimizu, H. and Urushihara, H.
TITLE	Sexual-cDNA in CSM
JOURNAL	Unpublished (1997)
COMMENT	Contact: Hideko Urushihara Institute of Biological Sciences University of Tsukuba 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan Email: da02huesakura.cc.tsukuba.ac.jp. Location/Qualifiers
FEATURES	

```

source
1. 638
/organism="Dictyostelium discoideum"
/strain="KAX3"
/db_xref="taxon:44689"
/clone="FC-AB21"
/clone_lib="Dictyostelium discoideum FC (H. Urushihara)"
BASE COUNT      247 a      128 g      203 t
BRIGINT

```

Query Match	1.4%	Score 38.8	DB 156	Length 658
Best Local Similarity	44.4%	Pred. No. 8.3		
Matches 157	Conservative	0	Mismatches 197	Indels 0
			Gaps	0
QY 1130	GTTCTGCTGCTGGTAGAGCCATCTACTTCCTATGATGCCATACACACAGAGATCTTCCACAA	1189		
Db				
Db 404	GTGTCGCTCCCTGGTTTAATTTGCCACATTTCTTCATCCATTTACCTACCATATCTGGCAATCT	345		
QY 1190	CAGTTACAGATGCTCTTAAGAAGTAAATGAGACTCCGGCAGATCTCTCAGTACATATATACAG	1249		
Db				
Db 344	CTGGATCTAAATGCACTACTACTGTTTGCATTTACTTTCACGAAATTTTCTTAATACCATGTGATG	285		
QY 1250	GGAACATCATCTTCACAGAGAGAAAAGTTATCAGACACAGAGCCGCACATTTCTAAAAATC	1309		
Db				
Db 284	GTAATGTCTCCCTCTTTCATTTATTCATATGATATCACTATACATAAATATCTGTTAAATC	225		
QY 1310	TTACTTCGAAGCTACTACAGCCTGTAACTCTTTTCAGAGAGCTCTATCTTTAAAAACATC	1369		
Db				
Db 224	TATCTGCTTTTACACTAATATTCATATCTGCTTCAACGTCTCTCTACATCAATATATAT	165		
QY 1370	GAGTGACCTCTGACAGACTCAGGCATTTCACTCAACAGGCGAGATTTCTGCTCGAAATGGAGC	1429		
Db				
Db 164	CAATAAATATCTCTCTCTTATTTGACATTTTTTTTTCTTTCTCTCTCTCTCTTTTAAAA	105		

Db 718 GAAATTAARTAAARAAAAAAMATWMAAAMGAAAGTAGTAGTAAGAAATAAAAAAMA 777

QY 1513 GCAAAATAGAACCAAGTACGTCAAAAAATCGACTTATTCGAAACATCACTTTA 1572  
 || ||| :| ||| ||| :| :| :| ||| |||

Db 778 WMAAGATTTAAACAAATATACCGTGCAGAAAAATTAACATTATTAACCGACTTTA 837

QY 1573 TTGGACCCGAGCGGACGCTTTATGAAATCATAGTTTAAAGAAATTCCTCAGCT 1625  
 || ||| :| ||| ||| :| :| :| ||| |||

Db 838 TTWCATAAAAAATWTATWTATGATATATGATATATWMAAAMAMAMCMAC 890

RESULT 9  
 CNS06D08/c 1067 bp DNA GSS 05-APR-2001  
 LOCUS T7 end of clone AR0AA009D07 of library AR0AA from strain CBS 732 of  
 DEFINITION Zygosaccharomyces rouxii, genomic survey sequence.

ACCESSION AL393198.1 GI:12143221  
 VERSION GSS; clone AR0AA009D07; RST AR0AA009D07CPI.  
 KEYWORDS Zygosaccharomyces rouxii.  
 SOURCE Zygosaccharomyces rouxii.  
 ORGANISM Zygosaccharomyces rouxii.  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.

REFERENCE 1 (bases 1 to 1067)  
 de Montigny,J., Straub,M., Potier,S., Tekala,F., Dujon,B.,  
 Wincker,P., Artiguenave,F. and Souciet,J.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 8.  
 JOURNAL Zygosaccharomyces rouxii  
 MEDLINE FEBS Lett. 487 (1), 52-55 (2000)  
 20584718

REFERENCE 2 (bases 1 to 1067)  
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
 Boloitin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,  
 de Montigny,J., Dujon,B., Durieux,P., Lepingle,A., Llorente,B.,  
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Wincker,P. and Weissenbach,J.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)  
 20584711

REFERENCE 3 (bases 1 to 1067)  
 Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAR-2001) Genoscope - Centre National de Sequencage,  
 MEDLINE 2 rue Gaston Creneau, CP 5706, 91057 Evry cedex, FRANCE. (E-mail:  
 20584711  
 3 (bases 1 to 1067)  
 segref@genoscope.cns.fr Web : www.genoscope.cns.fr)  
 COMMENT This BAC end sequence is part of a random genomic sequencing  
 program of thirteen yeast species:  
 . Saccharomyces bayanus var. uvarum,  
 . Saccharomyces exiguus,  
 . Saccharomyces servazzii,  
 . Zygosaccharomyces rouxii,  
 . Saccharomyces kluyveri,  
 . Kluyveromyces thermotolerans,  
 . Kluyveromyces lactis var. lactis,  
 . Kluyveromyces marxianus var. marxianus  
 . Pichia angusta,  
 . Debaryomyces hansenii var. hansenii,  
 . Pichia sorbitophila,  
 . Candida tropicalis and  
 . Yarrowia lipolytica.  
 Genomic inserts of 3 to 5 kb were prepared and both extremities  
 were sequenced. See keywords for description of this sequence and  
 for the sequence of the other extremity of this insert.

FEATURES  
 source  
 1. 1067  
 /organism="Zygosaccharomyces rouxii"  
 /strain="CBS 732"  
 /db\_xref="taxon:4956"  
 /clone="AR0AA009D07"  
 /clone\_lib="AR0AA"  
 /note="end : T7"

BASE COUNT 379 a 213 c 221 g 220 t 34 others  
 ORIGIN

Query Match 1.48; Score 39.2; DB 222; Length 1067;  
 Best Local Similarity 50.08; Pred. No. 7.6;  
 Matches 86; Conservative 4; Mismatches 82; Indels 0; Gaps 0;

QY 609 TGCCCTTACATTTGTGGAACCAAGGAGAGCTCTTTCTGACATTAATCTTCGGA 668  
 || ||| :| ||| ||| :| :| :| ||| |||

Db 471 TGGACACTACTGCTGCCGTACACAGCTACTTWTWTWTWTWTWTWTWTWTGCTGTTTGA 412

QY 669 TTCTGACCTGCATTTTATACAGAGCCTCGTGACTATTTCTATAATGCTAAAGTTTC 728  
 || ||| :| ||| ||| :| :| :| ||| |||

Db 411 TACTTCTACTGCTGCTTGTGACACTGCTGACACTTTTGTGTTATGACAGTTCG 352

QY 729 CTTTATTGACATTAAGTACAGAGCGAGCTCCTCAACAAGGGGAGTATAG 780  
 || ||| :| ||| ||| :| :| :| ||| |||

Db 351 CATGAGTCACACGTCGCGCTACTACAGCAATTCACCCCTTCGCAATATGTG 300

RESULT 10  
 A2680945 929 bp DNA GSS 14-DEC-2000  
 LOCUS EMTHQ31TR Entamoeba histolytica Sheared DNA Entamoeba histolytica  
 DEFINITION genomic, DNA sequence.

ACCESSION A2680945  
 VERSION A2680945.1 GI:11818091  
 KEYWORDS GSS.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Entamoeba histolytica.  
 Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 929)  
 Loftus,B., Van Aken,S. and Fraser,C.  
 TITLE Determination of clone end sequences from Entamoeba histolytica  
 JOURNAL HM1:IMSS shared DNA library  
 COMMENT Unpublished (2000)  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: bjoftus@tigr.org

FEATURES  
 source  
 1. 929  
 /organism="Entamoeba histolytica"  
 /strain="HM1:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHOSt1. Site 1: Bst I. Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark,  
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + 1 method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. In Genome  
 Sequencing: A Practical Approach, eds. M. Vaubin and B.  
 Bartell, Oxford University Press, 1999)."

BASE COUNT 313 a 146 c 102 g 368 t

```

source
1. 775
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3912241"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/organism="Homo sapiens"
/site_2="Salivary gland; Vector: PCMV-SPORE6; Site_1: NotI;
Average insert size 2.1 kb"
BASE COUNT 155 a 159 c 118 g 341 t 2 others
ORIGIN

Query Match 1.4%; Score 39.4; DB 141; Length 775;
Best Local Similarity 56.6%; Pred. No. 5.9;
Matches 73; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 2252 ATACGATCTGAAAACCAATATACATATCTACTGTTAAAGAGAGCTGGGGAATG 2311
|||||
DB 564 ATACGAGACTAGAGAAATGAAAGAACTTTGTTCTGTGTAAGAGATGGGGAATTT 505
|||||

QY 2312 ATAGTTTCCTTAAATTCGCTGGAAGAGCTCCATTTGCTTAATGAAAGTCTCTAT 2371
|||||
DB 504 TTGTATTCCTTAAAGAGCTTTTGAAGAGCTTTGTAATGTAAGTAAGTACTTTA 445
|||||

QY 2372 TTGACGAGT 2380
|||||
DB 444 GTTACCTGT 436

RESULT 7
LOCUS BF105345 952 bp mRNA EST 19-OCT-2000
DEFINITION 601822058R1 NIH_MGC_75 Homo sapiens CDNA clone IMAGE:4042310 3',
ACCESSION BF105345
VERSION BF105345
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10CM863 row: C column: 15
High quality sequence stop: 1.
Location/Qualifiers
1. 952
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4042310"
/clone_lib="NIH_MGC_75"
/lab_host="DH10B (T1 phage-resistant)"
/organism="Homo sapiens"
/site_2="Salivary gland; Vector: PCMV-SPORE6; Site_1: NotI;
Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones

```

```

BASE COUNT 245 a 222 c 231 g 252 t 2 others
ORIGIN

Query Match 1.4%; Score 39.4; DB 144; Length 952;
Best Local Similarity 50.8%; Pred. No. 6.4;
Matches 94; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1476 GGTCAATTAACATCAGTCTTATAGAGCGTCCAAAGAGCAAAAATGAAACCAAGCTAC 1535
|||||
DB 289 GTTAAATGATACCTGCTGTTGCTGCTAAAGAGGAGGATATATGACCTTAACACT 348
|||||

QY 1536 GTCAAAAATTCGACCTTATCTGAAACATCACTTATGAGCCGAGCGGCAAGCTTTTA 1595
|||||
DB 349 GTCAAAAATTCGACCTTATCTGAAACATCACTTATGAGCCGAGCGGCAAGCTTTTA 1595
|||||

QY 1596 TGAATAATCATGTTTAAAGAAATCTCAGTCATGACATCTTAAAGCTCAAGCTTCTGG 1655
|||||
DB 409 TGATACAGATGAACAGATATCAAGGAGTGCGACAGATGAAGGCGCAAGCTTTT 468
|||||

QY 1656 AACTG 1660
|||||
DB 469 ATCAG 473

RESULT 8
LOCUS CNS014SH 1064 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
ACCESSION BACN12916 of Drosophila library from Drosophila melanogaster (fruit
VERSION AL104555
KEYWORDS fly), genomic survey sequence.
SOURCE AL104555.1 GI:5616569
ORGANISM GSS.
REFERENCE fruit fly.
AUTHORS Plasmid Drosophila melanogaster
TITLE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
JOURNAL Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
COMMENT 1 (bases 1 to 1064)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billard at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
Location/Qualifiers
1. 1064
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN12916"
/organism="Drosophila melanogaster"
/site_2="Salivary gland; Vector: PCMV-SPORE6; Site_1: NotI;
Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones

```







117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 03:06:46 ; Search time 10300.6 Seconds  
(without alignments)  
2537.635 Million cell updates/sec

Title: US-09-446-677b-9  
Perfect score: 2787  
Sequence: 1 ATGAGAGTCCTTCCTCCCA.....GAGCAAAATACCAATCTCAA 2787

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*  
48: gb\_est48:\*  
49: gb\_est49:\*  
50: gb\_est50:\*  
51: gb\_est51:\*  
52: gb\_est52:\*  
53: gb\_est53:\*  
54: gb\_est54:\*  
55: gb\_est55:\*  
56: gb\_est56:\*  
57: gb\_est57:\*  
58: gb\_est58:\*  
59: gb\_est59:\*  
60: gb\_est60:\*  
61: gb\_est61:\*  
62: gb\_est62:\*  
63: gb\_est63:\*  
64: gb\_est64:\*  
65: gb\_est65:\*  
66: gb\_est66:\*  
67: gb\_est67:\*  
68: gb\_est68:\*  
69: gb\_est69:\*  
70: gb\_est70:\*  
71: gb\_est71:\*  
72: gb\_est72:\*  
73: gb\_est73:\*  
74: gb\_est74:\*  
75: gb\_est75:\*  
76: gb\_est76:\*  
77: gb\_est77:\*  
78: gb\_est78:\*  
79: gb\_est79:\*  
80: gb\_est80:\*  
81: gb\_est81:\*  
82: gb\_est82:\*  
83: gb\_est83:\*  
84: gb\_est84:\*  
85: gb\_est85:\*  
86: gb\_est86:\*  
87: gb\_est87:\*  
88: gb\_est88:\*  
89: gb\_est89:\*  
90: gb\_est90:\*  
91: gb\_est91:\*  
92: gb\_est92:\*  
93: gb\_est93:\*  
94: gb\_est94:\*  
95: gb\_est95:\*  
96: gb\_est96:\*  
97: gb\_est97:\*  
98: gb\_est98:\*  
99: gb\_est99:\*  
100: gb\_est100:\*  
101: gb\_est101:\*  
102: gb\_est102:\*  
103: gb\_est103:\*  
104: gb\_est104:\*  
105: gb\_est105:\*  
106: gb\_est106:\*  
107: gb\_est107:\*  
108: gb\_est108:\*  
109: gb\_est109:\*  
110: gb\_est110:\*  
111: gb\_est111:\*  
112: gb\_est112:\*  
113: gb\_est113:\*  
114: gb\_est114:\*  
115: gb\_est115:\*  
116: gb\_est116:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*

PR 01-DEC-1998; 98US-0110438.  
 XX (CONN-) CONNAUGHT LAB LTD.  
 XX Murdin AD, Oomen RP, Wang J;  
 PI WPI: 2000-412339/35.  
 DR N-PSDB; AAA30853, AAA30854.  
 XX  
 PT Nucleic acids encoding polypeptide antigens from Chlamydia useful for  
 PT preventing, diagnosing and treating diseases such as community acquired  
 PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset  
 PT asthma -  
 XX  
 PS Claim 16; Fig 7; 17app; English.

CC This sequence is a Chlamydia antigen of the invention, designated  
 CC CPN100639. The nucleic acids (and their complementary sequences) may be  
 CC used as diagnostic agents for detecting the presence of nucleic acids  
 CC encoding Chlamydia antigens in samples according to standard methods,  
 CC and therefore, for diagnosing Chlamydia infections. For example, they may  
 CC be used as primers and probes for diagnostic polymerase chain reaction  
 CC (PCR) assays. Antisense sequences may be used to down regulate  
 CC expression of the proteins and may be used to treat infections. The  
 CC nucleic acids may also be used to produce the protein antigens they  
 CC encode according to standard recombinant DNA methodologies. The  
 CC proteins may then be used as antigens for the production of antibodies  
 CC (i.e. as vaccines) for preventing infection by Chlamydia. The  
 CC antibodies may also be used as diagnostic reagents for detecting  
 CC infections. Chlamydia is a pathogen implicated in the development of  
 CC (for example) community acquired pneumonia, upper respiratory tract  
 CC disease (especially bronchitis and sinusitis, asthmatic bronchitis,  
 CC adult-onset asthma and acute exacerbations of asthma in adults.

XX Sequence 930 AA:

Query Match 40.2%; Score 1919; DB 21; Length 930;  
 Best Local Similarity 43.2%; Pred. No. 4.8e-118;  
 Matches 412; Conservative 164; Mismatches 329; Indels 48; Gaps 17;

QY 1 MKSSPKEVFTSTAIPLSH-IAT---ETVLDSSASFDGKNGKNGFSVRESQEDAGTTTLF 56  
 DB 1 mkp1pk11stclvrlp11stlatygadas1p1dsgagstf1pkstadangtyn1 60  
 QY 57 KGNVTLENTPGCTATITKSCFNFTKGD1FTGNGNSLLFOTVDAGTVAGAAVNSVVDKS 116  
 DB 61 sgnvyl-ndagkgtaltgcctctcgtlctgltfkgysfntvtaganagaa-stladka 118  
 QY 117 TTFIFGSSLSFISPGSSITTTGKAVSCSTGSLKFDKNVSLFSKNFSTD---NGAATA 173  
 DB 119 lltgfsnslftlaqpttrvasgkstls-sagalnldngltllfsqvnseannngaitc 177  
 QY 174 KTLSTLGTWMSALFSENTSSKKGAIQTSDALITITNGGEVFSNDTSSDSCAIFTTAS 233  
 DB 178 ktlisgnstslftltsnakk1gga1ysaaas1sgntq1vfmnkgelgga1gfaas 237  
 QY 234 VTSNNAKYSFDINKYTGASSTTGMSCGATCAVYKTSSTDTKTLINGNOMLFSNNTSTT 293  
 DB 238 ssitgnssj1fsgntatdaag-----kga1yckctgetp1t1sgnks1ltaensvt 291  
 QY 294 AGGAIYVKLELASGGLTFSRNSVNGTAPKGAIAIEDSGELSLADSGDIVFLGNTV 353  
 DB 292 qggalcabgldisaagp1t1fsmrcnctaagkga1aiads1s1s1s1s1s1s1s1s1s1 351  
 QY 354 TSTT-PTNNSSIDLSTSAKMTALRSAGRAITFYDPI---TTGSSTYTDVLAQVETPA 409  
 DB 352 tscaptstna1y1sgsak1tln1raagqsy1fydp1sant1gaa---dv1t1ngpds 407  
 QY 410 DSALQYTGNIITFGEKLSFTEADSKNLTSLQPTVLSGGTSLKHGVTLOQAFQOQA 469  
 DB 408 nspldysgt1v1sgk1sadaakaadn1f1k1qp1a1asg1a1kgnvel1dvngftqce 467

QY 470 DSRLENDVGTLEPADTSTIN--NLVINISSIDGAKKAKIETKATSKNLTSLGTTLLDP 527  
 DB 468 gstllmgpuk1k-ade1eas1t1k1v1d1s1a1eg1n1k1sv1e1t1e1g1n1k1t1l1s1p1f1q1d1s 526  
 QY 528 TGTVEHNSLRNPOSYDILELKAS-----GVTSTAVPDPPIMGKFRHYGOGTGW 578  
 DB 527 sgnfyshtingq1t1q1lv1vftaataas1y1d1a1l1t1sp1vq1r1p1e1-----hy1yq1g1h1w1e 580  
 QY 579 PIYWGCASTTATFNNTKGYIRNPERIGSLVPSNLSMNAFIDISSLHYLMETANEGLQD 638  
 DB 581 atwad1stak1sg1um1tw1t1q1p1n1p1er1ras1v1p1d1s1was1fd1r1l1q1m1ts1q1ans1y1q 640  
 QY 639 RAFPWAGLSNFEFKDSTKTRRCGRFRLHSGGVIGGNLHTCSDKILISAFCOLFGRDRDYE 698  
 DB 641 rgl1was1gtan1f1h1k1d1sg1c1n1ga1f1rh1ks1y1g1v1s1a1s1e1f1e1n1f1s1v1a1c1q1f1k1d1d1f1 700  
 QY 699 AKNGTIVYGGLTYOHNETYISLPCKLKPC--SLSYVPEIPIVLESGNLSYHTDNDLKT 756  
 DB 701 ventshnylas1y1q1h1ra1f1g1p1---mps1fg1st1d1m1k1d1p1l1l1na1q1s1y1t1k1nd1mt 757  
 QY 757 KYTTYPTVWGSWGNDSFALFEGRAPICL-DESALFEQVMEPMKIQFYVAHOBGFEQGT 815  
 DB 758 r1ys1p1e1a1g1s1w1c1n1s1g1a1e1l1g1s1a1y1l1p1k1e1a1f1g1y1f1p1l1k1q1a1v1s1r1q1n1f1e1s1g1a 817  
 QY 816 EAREFSSRLVNLALPIGIRFDEKSDCOQATVNLITGYVDLVRSPDPDCTTTRISGDSW 875  
 DB 818 e1a1r1a1f1d1g1d1v1n1s1l1p1v1g1l1r1e1k1s1e1d1e1k1n1f1e1s1l1a1y1g1d1y1r1k1p1r1s1t1m1v1s1g1a1s1w 877  
 QY 876 KYFGTNLARQALVLRAGNHPCFNSNFEAFSQQSFELRGSSRNYNVDLAKYOF 928  
 DB 878 ts1ck1n1a1r1a1q1a1f1a1s1a1g1s1h1l1s1p1h1e1l1s1g1e1a1e1y1e1r1s1a1h1y1n1v1d1c1g1y1s1f 930

Search completed: October 2, 2001, 03:26:53  
 Job time: 30245 sec









CC antibodies may also be used as diagnostic reagents for detecting  
 CC infections. Chlamydia is a pathogen implicated in the development of  
 CC (for example) community acquired pneumonia, upper respiratory tract  
 CC disease (especially bronchitis and sinusitis, asthmatic bronchitis,  
 CC adult-onset asthma and acute exacerbations of asthma in adults.

XX Sequence 885 AA;

Query Match 41.8%; Score 1998.5; DB 21; Length 885;  
 Best Local Similarity 47.2%; Pred. No. 2.6e-123;  
 Matches 422; Conservative 135; Mismatches 301; Indels 37; Gaps 16;

```

QY 51 GTTFLFGNLTLENIPGTAATSCFNNKRGDITFTGNGSLFQVYADACTVAGAANS 110
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 11 gldtcltclgtlclgnl gsaaltkqcsdtesistagkqysfslhks -saegaa-s 67

QY 111 SVVKSSTTFIOGSSLSFSPSSITT--GKAVSCSTGSLKPKNVSLFSKNFSTDN 168
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 68 vtcdknlslgfsstllaapsvlttprsgavkcg-gdlftfngtlllkqdyceeng 126

QY 169 GATTAKTLSLTGTMSALFSENISS---KKGAIQTSALTTTNOGEVFSQSDTSS 225
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 127 galstklslknsqtsfeqnksatqkkgalcatqvtlntlapcllfsmnaeag 186

QY 226 AAIPTFASVTSNNAKYSPIDNKVTGASSSTGMSGALCAVKTSDTKVTLNGNML 285
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187 galstgncitlgtstlvsensvt---alag--nggal-----sgdaavtlsgngv 235

QY 286 FSNNTSTTGAAGAVYKLELAS--GGLTFSRNSVNGTAPKGAIAIEDSGELSLADS 343
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 236 fsgnqvaangalaiyakkltlasggsggnpfennlvgtltaengaisllaegcslf 295

QY 344 GDYFLGNTYVSTTP-GTNSSIDLGTSAKMTALRSAGRAIYVDPITTTGSSITVD 402
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 296 gdhylngnaiatpqtktknsidigtgkheiralsghsiffydpitanaadstl 355

QY 403 KVNTPADSALQVYGNITFGEKLESETEADSKNLTSKLOPVTLSLKGVTLOT 462
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 356 nlmaadagnstqsgstlvsqklsedeakvaadltlklpvtltagnlvklkgvldt 415

QY 463 QAFQOADSRLMDVGTLEPA-DTSTINLVINISSIDGAKKAKIETKATSKNLTS 521
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 416 kgftqtagssvldagtlkasteevcltqlsdpvdsllgckkvlaasaaakvalsg 475

QY 522 ITLLDPGTREHNSLKNPQSYDLELKASCTVSTAVTPPIGGEFHYGQGTWPIV 581
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 476 lllldngnaeyendlgktqdfsfvqalsqatcltdvpavrpbyhygqtlw-mt 534

QY 582 W-----GTGASTATPFMMTKGYTPNPERIGSLVPSNLSMAAFIDISSLHYMETAN 636
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 535 wvddastpkrctetktlwtntgyipnpergplvpsnlwgsfslgqlgyieraltlc 594

QY 637 GDRAFWCAGLSNFHKDSTTRRGFRHLGGYVIGNLHTCSDKILSAFCQLGRDRDY 696
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 595 sdgfiwaayanfiddkkgckgkryrhksggyalgsaqaacsennlfafeqlgskdf 654

QY 697 FVAKNGTIVGGTIYVHNETHYISLPCKLRPCSLIYV---TEIPVLFSGULSTH 753
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 655 lvaknhdcltyagaflyqh---llecsgfigclldkllpagswshpvlvlegqlayshvnd 710

QY 754 LKTKYTYTPVYKSGMGSFALFEGRAPICIDESALFEQYMPMKQFYVAHGEFK 813
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 711 lktctyaypevkgsgwnnaimmjgaashypeylhctdclpapyikllylqdsfsek 770

QY 814 GTEAREGSSRLVIALPIGIRFDKESDCODATYNTLTGTVVDRNPDPCTTLRL 873
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 771 tgegrstddnlnfslspgkfkfcdndfysdlclslvpydlirndpactaivls 830

QY 874 SKTTFGTLNARQALVLRAGNHFCFNSFEARSGSEFLRGSSRNRYNDLGA 928
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 831 swetyannlarqalqvragshyafslpmfevlqgfivfvgssrlynvdlgskt 885
  
```

RESULT 10  
 AAW8423  
 ID AAW8423 standard; Protein: 928 AA.  
 AC AAW8423;  
 XX

XX 26-APR-1999 (first entry)

XX Chlamydia pneumoniae surface exposed protein Omp10.

XX Omp10; outer membrane protein 10; surface exposed protein;  
 KW antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.

XX Chlamydia pneumoniae.

XX WO9858953-A2.

XX 30-DEC-1998.

XX 19-JUN-1998; 98MO-DK00266.

XX 23-JUN-1997; 97DK-0000744.

XX (BIRK/) BIRKELUND S.

XX (CHR1/) CHRISTIANSEN G.

XX Birkelund S, Christiansen G, Knudsen K, Madsen A;  
 PI Mybind P;

XX WPI: 1999-105610/09.

XX N-PSDB; AAX06822.

XX Species-specific test for identifying mammals infected with

PT Chlamydia pneumoniae - comprises detecting antibodies specific for

PT outer membrane proteins of C. pneumoniae or nucleic acids encoding

PT these proteins

PS Claim 7; Page 60-62; 115pp; English.

XX This polypeptide comprises the novel 98.4 kDa surface exposed

CC protein Omp10 of the human respiratory pathogen Chlamydia

CC pneumoniae. Its amino acid sequence was deduced from DNA (see

CC AAX06822) isolated from a C. pneumoniae expression library. The

CC invention provides 12 novel surface exposed proteins. Omp4-Omp15

CC (see AAW8417-28), and nucleic acid sequences encoding them (see

CC AAX0616-27). A new species specific test is claimed that is used

CC to identify mammals (including humans) infected with Chlamydia

CC pneumoniae. The test comprises detecting antibodies specific for

CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer

CC membrane proteins, especially by PCR. The proteins are also used

CC in the diagnosis of C. pneumoniae infection in mammals. The

CC nucleic acids and proteins can also be used in the immunization of

CC mammals, the nucleic acids being particularly useful as DNA

CC vaccines for effecting in vivo expression of antigens. The

CC which are possibly associated with C. pneumoniae.

XX Sequence 928 AA;

Query Match 41.5%; Score 1981; DB 20; Length 928;  
 Best Local Similarity 45.2%; Pred. No. 3.9e-122;  
 Matches 429; Conservative 148; Mismatches 329; Indels 44; Gaps 19;

```

QY 1 MKSSPFKVFSTFAIFPLSM-----IATEVTLDSSASFDGNKGNFSVRSQDA-GTT 53
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 mksslhvfvlsslalpslfsafaaevneinlpntsfsg--pqtlytppeqttnadgtl 58

QY 54 YLEKGNVTLENIPGTGATIKSCFNNTKGDITFTGNGSLFQVYADACTVAGA 113
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 59 ynlgtgvsitn-egspaltasckettgnlsfghyqfllqpidag--ancftntaa 115
  
```



outer membrane proteins of *C. pneumoniae* or nucleic acids encoding these proteins

Disclosure: Page 78-80; 115pp; English.

This polypeptide is described as a subsequence of the claimed novel surface exposed protein Omp5 (see AAM8418) of *Chlamydia pneumoniae*, a human respiratory pathogen. The invention provides a new species specific test for identifying mammals (including humans) infected with *C. pneumoniae*. The test comprises detecting antibodies specific for surface exposed proteins Omp4-Omp5 (see AAM8417-28) or detecting nucleic acid fragments encoding them (see AAM8418-27), especially by PCR. The proteins are also used in the diagnosis of *C. pneumoniae* infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with *C. pneumoniae*.

Sequence 914 AA:

Query Match 41.9%; Score 2004; DB 20; Length 914;  
Best Local Similarity 46.4%; Pred. No. 1.2e-123;  
Matches 434; Conservative 140; Mismatches 318; Indels 44; Gaps 20;

```

1 MKSPEKFEF-STFAIF---PLSMITEVLDSSASFDGNKN-GNSVRESOEDACTYLL 55
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1 mnsqfswlslstlactscstvfataenlgspsdldgstntgtylxtnt--tlgldyl 58
56 FGNVTLNIPETGTATRTKSCFNMTKGLDFTGNGSLFQTVADAGTGAANSSVVK 115
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
59 lggdltlqnl-gdsaaltkgcfstdteslfsagkyksflnlks-saegaal-svtlck 115
116 STFTGFSSSLFASGSSITR--GKAVSCSTGSLKFKKNVSLFSKFNSTDNGAITA 173
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
116 nsltgfsstlftaapsvyltppsghavkcg-gdlftfnnlflfkqdyceengaalst 174
174 KTLSLGTMTMSALEFSENTS---KKGAIQTSDALITFNGCFVSEFTSSDSCGAIFT 230
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
175 knslknsngslsfegnksatgkkgalcatgvtidntmptlflmhaaagaalns 234
231 EASVTSNNAAKVSFDNKVTGASSSTGTGMSGCAICAVYSTDYKVTLLGNQMLFSNNT 290
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
235 tgnctltgntslvifsenavt---atag--nggal-----sgdadvtlsgnsvtfsqg 283
291 STTAGCAITVKKLELAS--GGLTFPSRNVNGCTAPKGAIAEDSGELSLSDSDIYF 348
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
284 avanggaiaiyakkltlaaggggagisfsmalvgqltagngaisllaagccslaaegdlcf 343
349 LGNTVSTTP-GTNRSSIDLGTSAKWTALRSAGRAIYFDPTTSSSTVVDVLKVNFT 407
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
344 ngnavatcpqtkrnsldgstakltlraalsghsiffydpilantaaadctclnlnka 403
408 PADDSALQYTGNIITFGEKISFTEADSKMLTKLQPTVLSGGLSLKAGVTLQTAFTQ 467
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
404 degnsuivgsyivfsekglsedeakvadnltstkqpvcllagnvlkcyvcltdkgyfcg 463
468 QADSRLKEMVGTLEPA-DSTIINNVLVNISSIDGAKKAKIETKATSKMLTSGITLLD 526
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
464 tagssvymaagtlkasteevcltglsipvdsllgegkavlaasaaskhvalasgplllid 523
527 PGTGFENSLRNPOSYDLLEKASGTVSTAVTPPIGKEFHYGOGTGWPIYV---- 582
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
524 ngngayenldlgtqdfstvglsalgtatttvpavptacphthgygvtgw-mkwvadt 582
583 -GTGASTATFNNKTGTYPNPBRIGSLVNSLMAAFIDISSLHLYMETANGLQDRAE 641
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
583 asprkktaktatwntlmglylpnpgyrgplvpsnlwsgldsiglylqysaltlcsdrfj 642
642 WCAGLSNFPHKDKSTRRGFRHLHSGGYVIGNLMHCSDKILSAAPQLGRDRDYVANK 701
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 643 waagyaanfldkdkgekkrkyrhsggyaigaagtcsemlsfafqglsgskdlfvakn 702
QY 702 QGVTVGGTLYOHNETYISLPCKLRPCSLSYV---TEIPVLESGNLSYTHNDNDKTKY 758
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 703 hcdtyagafyiqh----llecsgflgclldklpqswshkpiylegqlasvshsndtkty 758
QY 759 TTYPTVKGSMGNDSEFALIEGGRAPICLDESALFEQYMPMKLOFYVAHOGFKEQGTGAR 818
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 759 taypevkgsgwnafnmmlgaasshsyeyljhcdtyayiklnlylirgdsfsektyegr 818
QY 819 EFGSSKLVMLAPICGRFPEKSDCDATYNLTIGYVDLVRSPPDCTTLRLISGDSWKRF 878
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 819 sfldsnllfnlslpdygkiefcdsfydlctlsydpdlirndpkcttalvisgawely 878
QY 879 GTNLARQALVLRGNHFCNSNPEAFSOPFELRGS 914
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 879 anlaraqalqvragshyatspmfevlvgfvevrgs 914

```

RESULT 9

AA90238  
ID AAY90238 standard; Protein; 885 AA.

AA90238;  
29-AUG-2000 (first entry)

DE Mature *Chlamydia* antigen CPN100635.

KW *Chlamydia* antigen; diagnosis; infection; community acquired pneumonia;  
therapy; upper respiratory tract disease; bronchitis; sinusitis;  
asthma; adult-onset asthma; acute exacerbations of asthma.

OS *Chlamydia pneumoniae*.

PN WO200032794-A2.

PD 08-JUN-2000.

PF 01-DEC-1999; 99WO-CA01147.

PR 01-DEC-1998; 98US-0110339.

PR 01-DEC-1998; 98US-0110340.

PR 01-DEC-1998; 98US-0110427.

PR 01-DEC-1998; 98US-0110428.

PR 01-DEC-1998; 98US-0110438.

PA (CONN-) CONNANUGHT LAB LTD.

PI Murdin AD, Oomen RP, Wang J;

DR WPI; 2000-412339/35.

DR N-PSDB; AAA30849, AAA30850.

PT Nucleic acids encoding polypeptide antigens from *Chlamydia* useful for preventing, diagnosing and treating diseases such as community acquired pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset asthma -

PS Claim 16; Fig 3; 174pp; English.

XX This sequence is a *Chlamydia* antigen of the invention, designated CPN100635. The nucleic acids (and their complementary sequences) may be used as diagnostic agents for detecting the presence of nucleic acids encoding *Chlamydia* antigens in samples according to standard methods, and therefore, for diagnosing *Chlamydia* infections. For example, they may be used as primers and probes for diagnostic polymerase chain reaction (PCR) assays. Antisense sequences may be used to down regulate expression of the proteins and may be used to treat infections. The nucleic acids may also be used to produce the protein antigens they encode according to standard recombinant DNA methodologies. The proteins may then be used as antigens for the production of antibodies (i.e. as vaccines) for preventing infection by *Chlamydia*. The

30-DEC-1998.  
19-JUN-1998; 98WO-DK00266.  
23-JUN-1997; 97DK-0000744.  
(BIRK/) BIRKELUND S.  
(CHR1/) CHRISTIANSEN G.  
Birkelund S, Christiansen G, Knudsen K, Madsen A;  
Mygdin P;  
WPI: 1999-105610/09.  
N-PSDB: AAX06821.  
Species-specific test for identifying mammals infected with  
Chlamydia pneumoniae - comprises detecting antibodies specific for  
outer membrane proteins of C. pneumoniae or nucleic acids encoding  
these proteins  
Claim 7; Page 56-58; 115pp; English.  
This polypeptide comprises the novel 96.7 kDa surface exposed  
protein Omp4 of the human respiratory pathogen Chlamydia  
pneumoniae. Its amino acid sequence was deduced from DNA (see  
AAX06821) isolated from a C. pneumoniae expression library. The  
invention provides 12 novel surface exposed proteins, Omp4-Omp15  
(see AAW88417-28), and nucleic acid sequences encoding them (see  
AAX06816-77). A new species specific test is claimed that is used  
to identify mammals (including humans) infected with Chlamydia  
pneumoniae. The test comprises detecting antibodies specific for  
Omp4-Omp15 or detecting nucleic acid fragments encoding these outer  
membrane proteins, especially by PCR. The proteins are also used  
in the diagnosis of C. pneumoniae infection in mammals. The  
nucleic acids and proteins can also be used in the immunization of  
mammals, the nucleic acids being particularly useful as DNA  
vaccines for effecting in vivo expression of antigens. The  
vaccines may also prevent atherosclerosis and bronchial asthma,  
which are possibly associated with C. pneumoniae.

[illegible]

```
Db      346 t-ustgspavstrmaiaiaasnakflinlratcrnkvlfydpit---ssgatdkslnkad 4001
QY      409 ADSALQYTGNIITFTEKLSIETEADSKNLTSLQLPYTLISGTLISLKNHYTLOATFTQQ 4686
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      402 agsqnttgyvlvifgekliseeelkphnlktctfgavelaalgalvldyvltvanltty 4616
QY      469 ADSRLMDVGTLLP-PADSTIIINLVINISSIDCAKKAKIETKATSKNNLTSGTILLDP 5272
        | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      462 egskvwmvggttfteasaegvclngaiindidldgtinkalikalataaskdvalsprlmida 5216
QY      528 TCTFEFNISLNRPQSYDLDLELKASTVTVAVTDPIIMKEGHFGYCGTNGPIW---GT 5848
        | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      522 gqnyvehnlisqqyfplrliesaqumtltpitpbtltnthbyagvt-qlivwddat 5600
QY      586 GASTAFATEFWTKGTGIPMPERIGSLVPNSLNAFLIDISLHLYMETNEGIJOGRAFWCA 6444
        | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      581 aktkaatlwtcktykhpmepzqgplvpmlslagsfvtvdvrsjlsmdrtsalssatlvws 6400
QY      645 GLSNFFHNDSTFKTRGCFHILSGGYIVTIGMHLTCSDKITLSAFCOLFGRDRDYFAVKNOGT 7044
        | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      641 giadtlhedqgnqnsyhsasayalggtftaseentfnafcgdlfygdhdhlyakhnth 7000
QY      705 YVGGLTYQH--NETYSLPCKAPCSLSTVPEPIEVLSSGNLSYTHNTDNLKTAKTTY 7616
        | | | : | | | : : | : | | | | | | | | | | | | | | | | | | | |
Db      701 vyagamsyrhlheskltlaki-----lsgndsdlpfvinarfayghndnmktky 7520
QY      762 PTVKSNMDSNALPEFRGRAPICLD-ESALEEQVMPEMKILOFYVAHOEGFEQOSTEREF 8200
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      753 spvkgswnndaiglecggaipvvasgrtsvwdtlptlinemlayahndikengtegrsf 8120
QY      821 GSSRLVINALPLGIFNFDESQCOPATYNLTLGVVDLVRSNPDDTTTLRLISGDSMWKFTG 8800
        | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      813 qsecdflnalapvyglktefsd--kstyalstalaypdrilmdpgcettllmagsdwscgc 8700
QY      881 NIARQALYLRAHGNEFCENSNFENAEFSQSSEFULRGSSRNMYNDLGAKEYF 928
        | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      871 slsrdaalivragnhaafasnfevfsgqfvevelrsgysraialdgrriff 918
```

RESULT	9
AAW88429	
ID	AAW88429 standard; Protein; 914 AA.
XX	
AC	AAW88429;
XX	
DT	26-APR-1999 (first entry)
XX	
DE	Chlamydia pneumoniae surface exposed protein Omp5.
XX	
OS	Omp5; outer membrane protein 5; surface exposed protein; antigen;
KW	Infection; diagnosis; vaccine; atherosclerosis; asthma.
XX	
OS	Chlamydia pneumoniae.
XX	
PN	W09858953-A2.
XX	
PD	30-DEC-1998.
XX	
PF	19-JUN-1998; 98W0-DK00266.
XX	
PR	23-JUN-1997; 97DK-0000744.
XX	
PA	(BIRK/) BIRKELUND S.
XX	
PA	(CHRI/) CHRISTIANSEN G.
XX	
PI	Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI	Mygind P;
XX	
DR	WPI: 1999-105610/09.
XX	
DR	N-PSDB: AAX06828.
XX	

Species-specific test for identifying mammals infected with Chlamydia pneumoniae - comprises detecting antibodies specific for

```

RESULT 6
AA69369
ID AAY69369 strand: Protein; 918 AA.
AC AAY69369;
DE 19-JUN-2000 (first entry)
XX Amino acid sequence of the CPN100395 polypeptide.
DE CPN100395; Chlamydia infection; Immune response; vaccine.
XX Chlamydia pneumoniae.
XX WO200011183-A2.
XX 02-MAR-2000.
XX 18-AUG-1999; 99WO-IB01449.
XX 20-AUG-1998; 98US-0097187.
XX 20-AUG-1998; 98US-0097186.
XX 20-AUG-1998; 98US-0097189.
XX 20-AUG-1998; 98US-0097190.
XX 20-AUG-1998; 98US-0097195.
XX 20-AUG-1998; 98US-0097196.
XX 20-AUG-1998; 98US-0097197.
XX 27-AUG-1998; 98US-0097191.
XX 17-AUG-1999; 99US-0376770.
XX (CONN-) CONNAUGHT LAB LTD.
XX Murdin AD, Oomen RP;
XX WPI: 2000-224703/19.
XX N-PSDB; AAZ61509.
XX Novel antigens and corresponding DNA molecules that can be used to
XX prevent, treat and diagnose disease caused by Chlamydia infection in
XX mammals, especially humans -
XX Claim 19; Fig 15-E; 20PP; English.
XX AAY69362-69 represent Chlamydia pneumoniae polypeptides. The
XX polypeptides are present in the bacteriophage structure, in the
XX external vicinity of the membrane structure, in the inclusion membrane
XX structure, in the external vicinity of the inclusion membrane
XX and in the cytoplasm of the infected cell. The polypeptides may be
XX used to prevent, treat and detect the presence of Chlamydia infection
XX and/or the presence of Chlamydia in a sample. The polypeptides may
XX also be used to induce an immune response in a mammal. The vaccine
XX vector comprising the polynucleotides is used to induce an immune
XX response in a mammal. Antibodies directed against the polypeptides
XX may also be used therapeutically to treat and/or prevent a Chlamydia
XX infection.
XX Sequence 918 AA.
SQ

```

```

Query Match 42.2%; Score 2015; DB 21; Length 918;
Best Local Similarity 45.7%; Pred. No. 2, 2e-124;
Matches 433; Conservative 151; Mismatches 314; Indels 50; Gaps 20;

```

```

QY 1 MKSSPFKFESEFAIRPEPLMI---ATEVVLDSASFQDN-KNGNSVSESGEDA-GTYY 54
   |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1 mssstlillisslaipllmsvsadaadltlgsrdyngdstetltpaaasdaagty 60
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 55 LFKGNVTLLENIPGTGTAIRKSCFNNTKGDLTFTNGNSLFTQVVDAGTYAAGAVNSVVD 114
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 61 lldgvtysiq-agkqstlscstntagnltftgngfshfnalststvagvvnstlaas 119
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 115 KSTTFIGFSSLSFIASPGSSITTGKGAVSCSTGSLFKDKNVSILFESKNPSTDNGAIFAK 174
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

DB 120 gltkfsgfslclmlaer---ltgkgaikltdg-lvfesignldlneassengagaintk 175
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 175 TSLNCTGTMALFESSENTSSKKGAIQTSALRTTNGOGSEFSDNTPSSGAIATFEASV 234
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 176 flstlgsftrfvalflngssggqgaIyagsdvsisenagilstlgnmslsgaisaegn 235
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 235 TISNNAKVSFIDNKKVTGASSSTTGDMSGAICAYK--TSTDTVTTLTGNOMLFSNNTST 292
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 236 vlsnngnifldgckat-----engaidcnkengampdplltlsgneslflnntng 286
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 293 TAGGAIYVKKLELASG-GLTFESRNSVNGTAPKGAIAIEDSGEISLSDSDIYFLGN 351
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 287 nsqgaIytkllvlsrggyvlfnknkaanat-pkgaIaIldgeislsladnllifegn 345
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 352 TYTST---PGTNRSSIDLGTSKMTALNSAGRATFYDPRITGSSSTYTDVLAKNERP 408
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 346 t-tstgspasvtrnaiddlaskanajlnlratrgnkvlfydpl---ssgaIdklslnkad 401
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 409 ADSALQYTGNIITFTGKELSTEADSKNLTSKLLQPVTLSSGTLSSKHGVTLOTQAFTOQ 468
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 402 agsgnlyegyivfsgeliseelkpkdnkfstftgvelaagallvkdyvtvanlity 461
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 469 ADSRLMDYGTILE-PADSTINNLVINISSIDAKKAKIETKATSKNLTLSGTTILLDP 527
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 462 egskvmdggtlfeasaegytlnglalnldslgtnkalkakaakdvalsgplmIvda 521
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 528 TGFVFNHSLRNPOSYDILELKASGTVTSFAYVPDDIMGKFFHYGOGTGPVW---GT 584
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 522 qgnlyehnlsgqvfpilslsgtmtctldpdrplntlnlygyqgnwn-lwvddat 580
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 585 GASTATFNMWTKTYGILPNERIGSLVPNSLWNAFIDISSLHYLMETANEGLQDRAFWCA 644
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 581 aktnatlcwtktgrykpnpergplvpsnlsgfvdvrsqslmdstslsslnlws 640
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 645 GLSNFPHKSTKTRRGFRHLSGTYVIGNLHTCSDKILSAFQLGRDRDYVAANKGT 704
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 641 gladiIhedqkgnqrsyrsagyalggfflasenfifaqcqllgyddhIvakhth 700
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 705 VYGGTLYYOH---NERYISLPCKLRPCSLSYVPELIPVLPFSNMSTHPTNDKTYTYY 761
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 701 vyagamsyrlngesklakl-----lsgnadsIpfvtnarfayghdnmhtlyty 752
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 762 PTVKSGWGNDSFALBERGGRAPICLD-ESALFEQYMPFMKLOFYAHQEGKEQTEAREF 820
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 753 spkxsgwgnadafglecgallpvaasgrswdthpflnlemIyahqndtkengtegrsf 812
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 821 GSSRLVNLALPICIRFDKESDCODATYNTLGYTVDLVRSNPDCITTLKRISGSMKTFGT 880
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 813 gseelIflavpvgIkfeKfcd--kstydlslayvpydvirndpcttlmwsqswatcgt 870
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 881 NLAROALVIRAGNHPFNSFEAFSOFPELRGSSSNYNYNDLGAKXQF 928
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 871 slsrgaIlvragnhhaafsnfefsIqfvelIygsrsysalldgrrtgf 918
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 7
AA68422
ID AAM8422 standard; Protein; 918 AA.

```

```

XX AAM8422;
XX 26-APR-1999 (first entry)
DE Chlamydia pneumoniae surface exposed protein omp9.
XX Omp9; outer membrane protein 9; surface exposed protein; antigen;
XX infection; diagnosis; vaccine; atherosclerosis; asthma.
XX Chlamydia pneumoniae.
XX WO9858953-A2.
XX

```

QY 819 EFGSSRLVNLALPIGIRPEKSDCODATYNTLTGYVDLVRSNPDCTTLRISGDSKTF 878  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 819 sfddsnlfnslpvgkfkscndsfydltlsvpdlilnbpkctalaivsgaswely 878  
QY 879 GTNLARQALVLRAGNHCFNSNFESQSFELRGSSRNRYNDGAKYQF 928  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 879 annlarqalvragshyafspmfvlgqfvfeyrgsrllyvndlgkfgf 928

RESULT 5  
AAV90237 standard; Protein: 928 AA.  
ID AAV90237  
AC AAV90237  
DE 29-AUG-2000 (first entry)  
XX Chlamydia antigen CPN100635.  
XX Chlamydia antigen: diagnosis; infection: community acquired pneumonia;  
KM therapy: upper respiratory tract disease; bronchitis; sinusitis;  
KM asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma.  
XX Chlamydia pneumoniae.  
OS  
FH Key Location/Qualifiers  
FT Peptide 1..43  
FT /note= "signal peptide"  
FT Protein 44..928  
FT /note= "mature CPN100635"  
PN WC200032794-A2.  
XX 08-JUN-2000.  
XX 01-DEC-1999; 99WC-CA01147.  
XX 01-DEC-1998; 98US-0110339.  
PR 01-DEC-1998; 98US-0110340.  
PR 01-DEC-1998; 98US-0110427.  
PR 01-DEC-1998; 98US-0110428.  
PR 01-DEC-1998; 98US-0110438.  
PA (CONN-) CONNACHT LAB LTD.  
XX Murdin AD, Oomen RP, Wang J;  
PI WPI; 2000-412339/35.  
DR N-PSDB; AAA30849, AAA30850.  
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for  
PT preventing, diagnosing and treating diseases such as community acquired  
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset  
PT asthma.  
XX Claim 16; Fig 3; 174pp: English.

This sequence is a Chlamydia antigen of the invention, designated  
CPN100635. The nucleic acids (and their complementary sequences) may be  
used as diagnostic agents for detecting the presence of nucleic acids  
encoding Chlamydia antigens in samples according to standard methods,  
and therefore, for diagnosing Chlamydia infections. For example, they may  
be used as primers and probes for diagnostic polymerase chain reaction  
(PCR) assays. Antisense sequences may be used to down regulate  
expression of the proteins and may be used to treat infections. The  
nucleic acids may also be used to produce the protein antigens they  
encode according to standard recombinant DNA methodologies. The  
proteins may then be used as antigens for the production of antibodies  
(i.e. as vaccines) for preventing infection by Chlamydia. The  
antibodies may also be used as diagnostic reagents for detecting  
infections. Chlamydia is a pathogen implicated in the development of  
disease (especially bronchitis and sinusitis, asthmatic bronchitis,

CC adult-onset asthma and acute exacerbations of asthma in adults.  
XX  
SQ Sequence 928 AA;  
Query Match 42.5%; Score 2032; DB 21; Length 928;  
Best Local Similarity 46.3%; Pred. No. 1,76-125;  
Matches 440; Conservative 141; Mismatches 325; Indels 44; Gaps 20;  
QY 1 MKSFPKFFV-STPAIF--PLSMATETVLDSSAFPGKNK-GNFSRESQEDAGTYL 55  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 1 mksfswlvsstlaclscsvtaeaenlgspsdsgntlytpknc--ctgldyt 58  
QY 56 FKGNVLENIPTGTATKSCFNNTKGDITFGNGNSLTFQTVAGTVAGAANNSSVYDK 115  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 59 ltditlqnl-gdsaalkgcfsdteslftagyslsfiniks-saeagal-svtdk 115  
QY 116 STPIGSSLSFIASPGSSITF--GKAVSCSTGSLKEDKNVSLFSKNFTDNGAITA 173  
: : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 116 nsltgfsllflaapsvlttspgkavkcg-glditdngllfkdyceengalst 174  
QY 174 KTLSLGTGMSALFSENTS--KKGAIQTSDALTTGNGEVSFSQNTSSDGAALT 230  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 175 knlskntsgsisfegknsatgkkgalcatgtvdltnnaptlfsnnlaeaagains 234  
QY 231 EASYTISNNAKVPSFIDNKVITGASSSTTGDMSGAICAKYKSTDPKVTTLGNOMLFSNT 290  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 235 tgnctltgntslvtsensvt---atag-nggal-----sgddvltsgqvstfsgnq 283  
QY 291 STTAGGAIVYKLEIAS--GGLTLFSRNSVNGTAPKGAIAIEDSGELSLADSGDIVF 348  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 284 avanggalayakkllclaagsgggnpfnslivgttaagngalslaagcsfiseagdhyl 343  
QY 349 LGNTVSTTP-GTNRSSIDLCTSAKMTALRSACRAIYFDPIITGSGSTVTVLYKNET 407  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 344 ngnaivattpqtkrnsldsgtgdhelaigshtfipyilantaaadctclnlnka 403  
QY 408 PADALQYTGNIITFGEKLTSEEAADSKNLTKILOPTVSGSLTKHGVTLOTOAFTQ 467  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 404 dagnstdysgsvltsgeklseadeakvaadlvtclqkplvlaagnlvtkryvcltdkgtg 463  
QY 468 QADSRLEMDVGTTELEPA-DTSTINNLVINISSIDGAKAKIETATSKNLTLGTTLLD 526  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 464 tagssvmdagtlkasteerltglslpvdslgeqkvvlaaasaasnvalsp11ld 523  
QY 527 PTGTFYENHSLRNPQSIDILEKASGTVSTAATPDPIMGEKFRHVGCTGPIYV--- 582  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 524 ngnaayenhdlgtqdfsfvqslsalgtattdvpavpvaupthrygygtwg-mtwvddt 582  
QY 583 -GTGASTTATFNMKTKGYIPNPERIGSLVPSLWNAPDILSSLYHLMETAREGLQGDRAF 641  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 583 asptkktatlawntlylpnperqgplvpnslwgsfdaigvleraalclcsdgtf 642  
QY 642 WCAGLSNPFHKSDFKTRGFRHLRSGVYIGNLHTCSDKILASACCOLFGDRDPYVAKN 701  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 643 waavyanfldkdkkgyekrkyrhksyyaigaaqtcenlslsficqifgskdftlrvakn 702  
QY 702 QGTVYGGTLVYQHNHETVLSLPCKLSPCSLSYV--TEIPVLFSGNLSYTHNDLTKY 758  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 703 htdyagaftyqhn----ltecsfgicllklpgswshkplvleqqlayshvndlktky 758  
QY 759 TTYTPVAGSMGNDSEFALEFGGRAPICDESLFEQYMPFKMLQVYVAHOEFKROGFEAR 818  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 759 tayvevgsywnafnmmlgasshnspeylnctltypaylklntlylrqdsfsekgtegr 818  
QY 819 EFGSSRLVNLALPIGIRPEKSDCODATYNTLTGYVDLVRSNPDCTTLRISGDSKTF 878  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 819 sfddsnlfnslpvgkfkscndsfydltlsvpdlilnbpkctalaivsgaswely 878  
QY 879 GTNLARQALVLRAGNHCFNSNFESQSFELRGSSRNRYNDGAKYQF 928  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 879 annlarqalvragshyafspmfvlgqfvfeyrgsrllyvndlgkfgf 928

**RESULT 4**

AAW88418  
ID AAW88418 standard; Protein: 928 AA.

XX AAW88418:  
DT 26-APR-1999 (first entry)  
DE Chlamydia pneumoniae surface exposed protein Omp5.  
XX  
KM Omp5: outer membrane protein 5; surface exposed protein; antigen;  
XX infection; diagnosis; vaccine; atherosclerosis; asthma.  
OS Chlamydia pneumoniae.  
XX  
XX W09858953-A2.  
PN 30-DEC-1998.  
PD 30-DEC-1998.  
PF 19-JUN-1998; 98WO-DK00266.  
XX  
PR 23-JUN-1997; 97DK-0000744.  
XX  
PA (BIRK/) BIRKELUND S.  
PA (CHR1/) CHRISTIANSEN G.  
XX  
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;  
PI Mygind P;  
XX  
DR MPI; 1999-105610/09.  
DR N-PADB; AAX06817.  
XX

PT Species-specific test for identifying mammals infected with  
PT Chlamydia pneumoniae - comprises detecting antibodies specific for  
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding  
PT these proteins  
PS  
PS Claim 7; Page 43-45; 11pp; English.

XX This polypeptide comprises the novel 97.2 kDa surface exposed  
CC protein Omp5 of the human respiratory pathogen Chlamydia  
CC pneumoniae. Its amino acid sequence was deduced from DNA (see  
CC AAX06817) isolated from a C. pneumoniae expression library. The  
CC invention provides 12 novel surface exposed proteins, Omp4-Omp5  
CC (see AAW88417-28), and nucleic acid sequences encoding them (see  
CC AAX06816-27). A new species specific test is claimed that is used  
CC to identify mammals (including humans) infected with Chlamydia

CC pneumoniae. The test comprises detecting antibodies specific for  
CC OmpA-omp5 or detecting nucleic acid fragments encoding these outer  
CC membrane proteins, especially by PCR. The proteins are also used  
CC in the diagnosis of *C. pneumoniae* infection in mammals. The  
CC nucleic acids and proteins can also be used in the immunization of  
CC mammals, the nucleic acids being particularly useful as DNA  
CC vaccines for effecting *in vivo* expression of antigens. The  
CC vaccines may also prevent atherosclerosis and bronchial asthma,  
CC which are possibly associated with *C. pneumoniae*.

**SQ Sequence 928 AA;**

Query Match 43.38; Score 2069; DB 20; Length 928;

best local similarity 40.58; freq. no. 0.36 120;  
Matches 446; Conservative 141; Mismatches 319; Indels 44; Gaps 20;

QY 1 MKSSFPKVF-STFAIF---PLSMIAETVLDDSSASFDGNKN-GNFSVRESQEDAGTYL 55

Db 1 mksqfswlvlsstlaclftscstvfafaataenlgpsdsfdgstntgtcypknt--ttgidyt 58

QY 56 FKGNVTLLENIPGTGTATKSCFNNTKGDLTFTGNGNSLLFQTVDA GTVAGAAVNSSVVDK 115

```

Db 59 lttqditlqnl-qdsaaltkqcfdsdtteslsfaqqkqyslsfniks-saeqaal-svttcdk 115

```

OV 116 STTFIGFSSLSFIASPCSSITT--GKAVSCSTGSLKFEDKNVSLFSKNFESTDNGAITA 173

[illegible]

OV 174 KTI.SITGTTMSAT.ESENTSS---KKGGA.TOTSDA.ITTGNOGEVSESDNTSS.DSGAATET 230

Db 175 kn|s|f|e|n|k|s|a|t|n|k|a|a|c|a|t|a|t|v|d|t|n|t|a|n|t|f|s|n|f|a|e|a|a|a|a|n|s 234

331 EASYTISNNAKVSETDNKYMCASSSTGCDMGGCAICAVKKTSTDTKYMTTCGNOMIESNNT 290

[illegible][illegible]

..

[illegible][illegible]

20 2019-2020

```

0Y      | :: ||::||::||| | : ||| | ||||: | ||||: |||
      408 FADSDALQ I GN LI E GENLSE I EADSNNLI SNLDQF V LSSG ILSNHSV I DQ I QAF IQ 409

```

DB 404 aagnstaysgsivisgekrsedeakvaadnltstlkqpvllagnlvikgvlalckgllq 403

QY 468 QADSRLEMDVGITLEPA-DISTINNLVINSSIDGAKKAKIEIKATSKNLTLSGITLED 326

Db 464 tagssvImdagttlikastevttlttgsIpvdsIgeggkvvIaasaasknvalsgpIId 523

QY 527 PTGTFEYENHSLRNPQSYDILELKASGTVTSTA VTPDPIMGEKEFHGYQGTWGP I W --- 582Z

Db 524 nqgnayenhdlgktqdfsfvqlsalgtatttdvpavptvatpthygygtwg-mtwvddt 582

QY 583 -GTGASTTATFNWTKTGYPNPERIGSLVPNSLWNAFIDISSLHYLMETANEGLOGDRAF 641

Db 583 astpkkatatlwtntgylpnpergplvpnslwgsfsdlaigvlersaltlcldrgf 642

QY 642 WCAGLSNFEHKDSTKTRGFRHLSGGYVIGGNLHTCSDKILSAFCQLFGRDRDYFAKN 701

Db 643 waagvanfldkdkkgekrkyrhksggyaiggaqtcsenl1sfafcqlfgsddflvkn 702

QY 702 QGTVYGGTLYQHNETYISLPCKLRPCSLSYV--TEIPVLESGNLSYTHTDNDLKY 758

Db 703 htdtyagafyiqh-----ltecsqfigclldklpqswhkplvlegqlaysbvsnldlktky 758

QY 759 TTYPTVKGSWGNDSFALEFGGRAPICLDESALFEQYMPFMKIQFYVAHQEGFKEQGTAR 818

Db 759 t a y e v k g s w g n a f n m l g a s s h s y p e y l h c f d t y a p y i k n l t y i r g d s f s e k t e g r 818

CC membrane protein can be used as a vaccine to provide protection against  
 CC Chlamydia infections, especially Chlamydia pneumoniae infections.  
 CC The present polypeptide may also be administered orally to treat  
 CC Chlamydia infection.

XX Sequence 928 AA;

Query Match 99.6%; Score 4760; DB 21; Length 928;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 925; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKSFPKVFSTFAIFPLSMATEVLELSDSSAFDGNKGNFVSRESCDATTYLFKGNV 60  
 DB 1 mkssfpkfvtfaifplsmatetvldssasfdgnknfnsvrescdattylfkgnv 60  
 QY 61 TLENIPTGTAITKSCFNNTKGDLPFTGNGNSLLFQTVDACTVAGAAVSSVVDKSTTFI 120  
 DB 61 tleniptgtaitkscfnntkgdlpftgngnsllefqvdaactvagaavssvvdksttffi 120  
 QY 121 GFSSLSFIASPGSSITTGKGVSCSTGSLKFKDKVNSLLFSKNESTDNGATTAATLSLTG 180  
 DB 121 gfsslsfiasspgssittgkgvscstgslskfkdkvnsllfsknestdngattaatlsltg 180  
 QY 181 TTMSALFSEMTSSKKGAIQTSDALTTGNGEVSFSDNTSSDGAATFTPASVTISNNA 240  
 DB 181 ttmsalfsemtsskkgaiqtstdalittgngevsfndntssdgaatftpasvtisnna 240  
 QY 241 KVSFIDNRKVTGASSSTTGDSGAIICAVKTSSTDVKTLTGQMLFSSNNTSTAGAIYV 300  
 DB 241 kvsfidnrkvtgasssttgdsgaiicavktsstdvktlvtgqmlfssnntstagaaiyv 300  
 QY 301 KKLELAGGGLTFERNVNGTAPKGAIAIEDGELSLSADSDIYFLGNTVSTTPGT 360  
 DB 301 kklelagggltfervnngtapkgaiaiedgelslsadsdiyflgntvsttpgt 360  
 QY 361 NRSIDIDGTSAKMTALRSAGRAIFYDPITTSSTVTDLAKNEFPADSALOYTGNI 420  
 DB 361 nrsididgtakmtalrsagraifydpitststvtldaknefpadsaloytgnii 420  
 QY 421 FTGEGKSETPADSKNLTJKLOPVLISGCTLKHKVTLQTOAFQOADSRLKEMDVGT 480  
 DB 421 ftgeksetpadsknltjklpvlisgctlshkvvtlqtqafqoadsrlemdvgt 480  
 QY 481 LEPADDTINNVLNINISIDGAKKAKTETKATSKNLTLSGITTLIDPGRFFYEHSLRN 540  
 DB 481 lepaddtinnvlninisidgakkaktetkatsknltlsgitllidpgrffeyehslrn 540  
 QY 541 QSYDILELKASGTVTSAVTPDPPIMGKFFHYGYOGTWGPIYWGTAATFNNTKTGYI 600  
 DB 541 qsydilelkasgvtvsaavtpdpimgkffhygyogtwgpiywgtaatfnnktgyi 600  
 QY 601 PNEPRISGLVPNSLIMAFIDISSLHYMETANBGLGDRAPWCAGLSNFFHKDSTKTRRG 660  
 DB 601 pnerisglvpsnlmafidisslhymetanbglgdrapwcaaglsnffhkdstktrrg 660  
 QY 661 FRHLSGGYVIGNTHRCSDRIISAFCOLGRDRDYFVANKOGVYGGCTLYYONENYIS 720  
 DB 661 frhlsogyvignthrcsdrisafcolgrdrdyfvankogvyggctlyyonenysis 720  
 QY 721 LPCKLRCSLSYVTEIPVLFSGMLSTHTDNDLKTFTYTPYVKGSGMDSFALFEGR 780  
 DB 721 lpcklrclsyvtelplvfsgmlsthtdndlktftypvyksgmddfalegr 780  
 QY 781 APICLDESALFEQYMPMKLQFYVAHQEGFEQGEAREREGSSRLVNLALPIGIRPKES 840  
 DB 781 apicldesalfeqympmklqfyvahqegfeggearefegssrlvnalalpigirpk 840  
 QY 841 DCQDATNLTGLTGYVLDVLRNSPCTTTLRTISGDSMKTFFGNLAQALVLRAGNFCNSN 900  
 DB 841 dcqdatnltgltyvldvlrnspecttllrtisgdsmtffgnlaqalvlragnfccnsn 900  
 QY 901 FEAFSOFSEFLRGSRRNYNVDLGAKYOF 928

DB 901 featsqsfelrgssrrnyvndlgaakyof 928

RESULT 3

AAV34611  
 ID AAV34611 standard; Protein; 597 AA.

AC AAV34611;  
 DT 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae transmembrane protein sequence.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN W09927105-A2.

PD 03-JUN-1999.

PE 20-NOV-1998; 98MO-1B01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

PA (GENST ) GENSET.

PI Griffiths R;

DR WPI: 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

PS Page 632-633; Disclosure; 1912pp; English.

XX AAV34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAV31990) of Chlamydia pneumoniae.  
 CC C. pneumoniae causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
 CC frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in  
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
 CC nucleotide sequences can also be used as immunogenic compositions,  
 CC especially where the vector directs the expression of a neutralising  
 CC epitope of C. pneumoniae.

XX Sequence 597 AA;

Query Match 58.9%; Score 2816.5; DB 20; Length 597;  
 Best Local Similarity 94.1%; Pred. No. 2e-177;  
 Matches 562; Conservative 6; Mismatches 18; Indels 11; Gaps 3;

QY 149 LKPKKNVSLFSKNESTDNGATTAATLSLTGTTMSALFSENTSSKKGAIQTSDALIT 208  
 DB 1 lfkknvslfsknestdngattaatlsiltgtmsalfsentskkgaiqtstdalit 208  
 QY 209 GNGEVSFSDNTSSDGAATFTASVTISNNAKVSFIDNKVTVGASSSTGDMSCAICAY 268  
 DB 61 gngevsfndntssdgaatftasvtilsnakvsfldnkvtvgassstgdmsgaicay 268  
 QY 269 KTSPTDRTVTTGNOMLFSNNTSTTAGAIYVKKLELASGLTFSRNSVNGTAPKGA 328  
 DB 121 ktsdrtvtltgnmlfssnntsttagaiyvkklelasgltfssrnsvngtapgga 328  
 QY 329 IATEDGELSLSADSGDIVLGNTVSTTPGTNRSSIDLQTSAKMTALRSAGRAIFYD 388  
 DB 181 iatedgelslsadsgdivlgntvsttpgtnrssidlqtsakmtalrsagraifyd 388

outer membrane proteins of *C. pneumoniae* or nucleic acids encoding these proteins

Claim 7; Page 53-55; 115pp; English.

This polypeptide comprises the novel 90.0 kDa surface exposed protein Omp8 of the human respiratory pathogen *Chlamydia pneumoniae*. Its amino acid sequence was deduced from DNA (see AA06820) isolated from a *C. pneumoniae* expression library. The invention provides 12 novel surface exposed proteins, Omp4-Omp15 (see AA06817-28), and nucleic acid sequences encoding them (see AA06816-27). A new species specific test is claimed that is used to identify mammals (including humans) infected with *Chlamydia pneumoniae*. The test comprises detecting antibodies specific for Omp4-Omp15 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used in the diagnosis of *C. pneumoniae* infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with *C. pneumoniae*.

Sequence 928 AA:

Query Match 100.0%; Score 4778; DB 20; Length 928;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 928: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKSPPKVFSTFAIPPLMTETVLDSASFQDNKNGFVSRSQEDAGTTLFKGNV 60  
1 mksppkfvfstfaipplmtetvldssafqdnknngfvsrsqedagtytlfkgnv 60

61 TLENPGTGATATKSCFNNTKDLPFTGNGSNLRFQYDAGVAAANVSVYDKSTFPI 120  
61 tlenpgtgatatkscfnnTKDLPFTGNGSNLRFQYDAGVAAANVSVYDKSTFPI 120

61 TLENPGTGATATKSCFNNTKDLPFTGNGSNLRFQYDAGVAAANVSVYDKSTFPI 120  
61 tlenpgtgatatkscfnnTKDLPFTGNGSNLRFQYDAGVAAANVSVYDKSTFPI 120

121 GFSSLSFIFSPGSSITTGGAIVSCSTGSLKFPDKNVSLFKNFSTPNDGCAITAKTLSTLG 180  
121 gfsslsfifspgssitTGGAIVSCSTGSLKFPDKNVSLFKNFSTPNDGCAITAKTLSTLG 180

121 GFSSLSFIFSPGSSITTGGAIVSCSTGSLKFPDKNVSLFKNFSTPNDGCAITAKTLSTLG 180  
121 gfsslsfifspgssitTGGAIVSCSTGSLKFPDKNVSLFKNFSTPNDGCAITAKTLSTLG 180

181 TTMALFSENTSSKKGALQTSALITFGNOGEVSFSDNTSSDGAIFTEASVTISNNA 240  
181 ttmalfseentsskkgalqtsalITFGNOGEVSFSDNTSSDGAIFTEASVTISNNA 240

181 TTMALFSENTSSKKGALQTSALITFGNOGEVSFSDNTSSDGAIFTEASVTISNNA 240  
181 ttmalfseentsskkgalqtsalITFGNOGEVSFSDNTSSDGAIFTEASVTISNNA 240

241 KSPFDNKVTGAGSSSTGDMGCAICAKRTSDTKVTLTGNGMLPFSNNTSTAGCAIIV 300  
241 kspfdnkvtgagssstGDMGCAICAKRTSDTKVTLTGNGMLPFSNNTSTAGCAIIV 300

241 KSPFDNKVTGAGSSSTGDMGCAICAKRTSDTKVTLTGNGMLPFSNNTSTAGCAIIV 300  
241 kspfdnkvtgagssstGDMGCAICAKRTSDTKVTLTGNGMLPFSNNTSTAGCAIIV 300

301 KLELASGGLTFERNVNGVGPAPKGAIAEDSGELSLADSGDIVFLGNTVSTTPTCT 360  
301 klelasgglTFERNVNGVGPAPKGAIAEDSGELSLADSGDIVFLGNTVSTTPTCT 360

301 KLELASGGLTFERNVNGVGPAPKGAIAEDSGELSLADSGDIVFLGNTVSTTPTCT 360  
301 klelasgglTFERNVNGVGPAPKGAIAEDSGELSLADSGDIVFLGNTVSTTPTCT 360

361 NRSIDLGTSAMKMLRSAGRAIYFDPITTGSSSTVTDLKVNETPADSALOYTGNI 420  
361 nrsidlgtsamkmlrsagraIYFDPITTGSSSTVTDLKVNETPADSALOYTGNI 420

361 NRSIDLGTSAMKMLRSAGRAIYFDPITTGSSSTVTDLKVNETPADSALOYTGNI 420  
361 nrsidlgtsamkmlrsagraIYFDPITTGSSSTVTDLKVNETPADSALOYTGNI 420

421 FPGKLSFSEADSKNLSKLOPYTLISGCTSLKRGVTLQQAATQOQDSLEMDVGT 480  
421 fpgklsfseadsknlskLOPYTLISGCTSLKRGVTLQQAATQOQDSLEMDVGT 480

421 FPGKLSFSEADSKNLSKLOPYTLISGCTSLKRGVTLQQAATQOQDSLEMDVGT 480  
421 fpgklsfseadsknlskLOPYTLISGCTSLKRGVTLQQAATQOQDSLEMDVGT 480

481 LEPADTSTNNLVINISSIDGAKKAKIETKATSKNLTLSGTTLLDPTGFEENSLRNP 540  
481 lepadtstnnlvINISSIDGAKKAKIETKATSKNLTLSGTTLLDPTGFEENSLRNP 540

481 LEPADTSTNNLVINISSIDGAKKAKIETKATSKNLTLSGTTLLDPTGFEENSLRNP 540  
481 lepadtstnnlvINISSIDGAKKAKIETKATSKNLTLSGTTLLDPTGFEENSLRNP 540

541 GSYDILKAKSGTVSTAVTPDPIMGEKPHGYOGTWGPVWGTAAGTATFNTKTGYI 600  
541 gsydILKAKSGTVSTAVTPDPIMGEKPHGYOGTWGPVWGTAAGTATFNTKTGYI 600

541 GSYDILKAKSGTVSTAVTPDPIMGEKPHGYOGTWGPVWGTAAGTATFNTKTGYI 600  
541 gsydILKAKSGTVSTAVTPDPIMGEKPHGYOGTWGPVWGTAAGTATFNTKTGYI 600

601 PUPERIGSLVPSINWAFDISSLHYMETANEGLOGDAFACAGISNFHHOSKTRRG 660  
601 puperIGSLVPSINWAFDISSLHYMETANEGLOGDAFACAGISNFHHOSKTRRG 660

601 PUPERIGSLVPSINWAFDISSLHYMETANEGLOGDAFACAGISNFHHOSKTRRG 660  
601 puperIGSLVPSINWAFDISSLHYMETANEGLOGDAFACAGISNFHHOSKTRRG 660

661 FPHLSGGYVIGNMTHTCSDKILISAAFCQLFGRDRDYFVAKNGVTYGGTLYYQHNETYIS 720  
661 fphlsGGYVIGNMTHTCSDKILISAAFCQLFGRDRDYFVAKNGVTYGGTLYYQHNETYIS 720

661 fphlsGGYVIGNMTHTCSDKILISAAFCQLFGRDRDYFVAKNGVTYGGTLYYQHNETYIS 720  
661 fphlsGGYVIGNMTHTCSDKILISAAFCQLFGRDRDYFVAKNGVTYGGTLYYQHNETYIS 720

721 LPCKLRPCLSVTPPEIYVYLSGNSLSTHTNDLTKTKTYYTVYVGSNGNSFALFEFG 780  
721 lpcklrpCLSVTPPEIYVYLSGNSLSTHTNDLTKTKTYYTVYVGSNGNSFALFEFG 780

721 LPCKLRPCLSVTPPEIYVYLSGNSLSTHTNDLTKTKTYYTVYVGSNGNSFALFEFG 780  
721 lpcklrpCLSVTPPEIYVYLSGNSLSTHTNDLTKTKTYYTVYVGSNGNSFALFEFG 780

781 APICDESALFQVYWPFKLOVYVHOGFROGTEAREFGSSRLVNLALPGIFEDKES 840  
781 apicDESALFQVYWPFKLOVYVHOGFROGTEAREFGSSRLVNLALPGIFEDKES 840

781 APICDESALFQVYWPFKLOVYVHOGFROGTEAREFGSSRLVNLALPGIFEDKES 840  
781 apicDESALFQVYWPFKLOVYVHOGFROGTEAREFGSSRLVNLALPGIFEDKES 840

841 DCODATYVNLITGYTVDLVRSNPDCTTTIRISGDSMKRTGTNLARQALVLRAGNHCFNSN 900  
841 dcoDATYVNLITGYTVDLVRSNPDCTTTIRISGDSMKRTGTNLARQALVLRAGNHCFNSN 900

841 DCODATYVNLITGYTVDLVRSNPDCTTTIRISGDSMKRTGTNLARQALVLRAGNHCFNSN 900  
841 dcoDATYVNLITGYTVDLVRSNPDCTTTIRISGDSMKRTGTNLARQALVLRAGNHCFNSN 900

901 FEAFSOFSEPLRGSSRNINVDIGAKYQF 928  
901 feafSOFSEPLRGSSRNINVDIGAKYQF 928

901 FEAFSOFSEPLRGSSRNINVDIGAKYQF 928  
901 feafSOFSEPLRGSSRNINVDIGAKYQF 928

RESULT 2

AAV94327 standard; Protein; 928 AA.

AAV94327;

11-AUG-2000 (first entry)

*Chlamydia pneumoniae* 98kD putative outer membrane protein.

*Chlamydia*; antigen; vaccine; infection; outer membrane protein.

*Chlamydia pneumoniae*.

WO2000026237-A2.

11-MAY-2000.

29-OCT-1999; 99WO-GB03579.

29-OCT-1998; 98US-0106070.

01-MAR-1999; 99US-0122066.

27-OCT-1999; 99US-0428122.

(CONN-) CONNAGHT LAB LTD.

Murdin AD, Oomen RP, Dunn PJ;

WPI; 2000-365569/31.

N-PSDB; AAA27021.

Novel *Chlamydia* 98 kDa putative outer membrane protein antigen, used for vaccination and protection against *Chlamydia* infection

Claim 6; Fig 1; 93pp; English.

The present sequence is the 98kDa putative outer membrane protein from *Chlamydia pneumoniae*. The genomic sequence was amplified using two PCR primers. The 5' primer contains a NotI restriction site, a ribosome binding site, an initiation codon and a sequence close to the 5' end of the 98kDa putative outer membrane protein coding sequence. The 3' primer contains the sequence encoding the C-terminal sequence of the putative outer membrane protein and a BamHI restriction site. The stop codon was excluded and an additional nucleotide was inserted to obtain an in-frame C-terminal fusion with the Histidine tag. The PCR product was cloned into a eukaryotic expression vector (pCA-Myc-His) by restricting both the vector and the PCR product with NotI and BamHI and performing a ligation reaction. This expression vector was injected intramuscularly and intranasally into mice, which were subsequently inoculated with *Chlamydia pneumoniae*. The chlamydial lung titers of the immunised mice were lower than those of the controls. Thus the 98kDa putative outer



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2001, 03:26:41 ; Search time 97.54 Seconds  
(without alignments)  
576.779 Million cell updates/sec

Title: US-09-446-677b-10

Perfect score: 4778  
Sequence: 1 MKSFPKVFSTFAIFPLSM.....FELRGSSRNRYNDLAKYQF 928

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_0601:\*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:\*
- 22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	4778	100.0	20	AAW88421
2	4760	99.6	21	AAV94327
3	2816.5	58.9	20	AAV34611
4	2069	43.3	20	AAW88418
5	2032	42.5	21	AAV90237
6	2015	42.2	21	AAV69369
7	2009	42.0	20	AAW88422
8	2004	41.9	20	AAW88429
9	1998.5	41.8	21	AAV90238
10	1981	41.5	20	AAW88423
11	1971	41.3	21	AAV90239

12	1951	40.8	936	21	AAV9842	Chlamydia pneumoni
13	1932.5	40.4	925	21	AAV9843	Chlamydia pneumoni
14	1929	40.4	930	20	AAV35052	Chlamydia pneumoni
15	1919	40.2	930	21	AAV90240	Chlamydia antigen
16	1916	40.1	930	20	AAW88424	Chlamydia pneumoni
17	1912.5	40.0	927	20	AAV35054	Chlamydia pneumoni
18	1854	38.8	949	20	AAV35060	Chlamydia pneumoni
19	1853	38.8	928	20	AAW88417	Chlamydia pneumoni
20	1853	38.8	945	21	AAV90236	Chlamydia antigen
21	1830	38.3	945	21	AAV69368	Antio acid sequenc
22	1809	37.9	945	20	AAW88428	Chlamydia pneumoni
23	1572.5	32.9	643	20	AAV35056	Chlamydia pneumoni
24	1559	32.6	841	21	AAV92818	C. pneumoniae CPN1
25	1553	32.5	841	20	AAW88420	Chlamydia pneumoni
26	1430.5	29.9	922	21	AAV95548	Chlamydia pneumoni
27	1423.5	29.8	922	20	AAW88419	Chlamydia pneumoni
28	1410.5	29.5	922	20	AAV34597	Chlamydia pneumoni
29	1377.5	28.8	973	21	AAV96274	Chlamydia POM91B
30	1245	26.1	671	20	AAV35050	Chlamydia pneumoni
31	1214	25.4	230	20	AAV34612	Chlamydia pneumoni
32	1203.5	25.2	507	20	AAV34614	Chlamydia pneumoni
33	1128	23.6	1132	20	AAV35048	Chlamydia pneumoni
34	1121	23.5	1012	20	AAV16735	C. trachomatis LGV
35	1117.5	23.4	982	21	AAV13633	C. trachomatis pmp
36	1112.5	23.3	1006	21	AAV13639	C. trachomatis pmp
37	1105.5	23.1	1013	20	AAV16737	C. trachomatis B s
38	1093.5	22.9	1013	20	AAV16738	C. trachomatis F s
39	1015.5	21.3	450	20	AAV34617	Chlamydia pneumoni
40	918	19.2	610	20	AAW88431	Chlamydia pneumoni
41	888	18.6	530	20	AAV35064	Chlamydia pneumoni
42	856	17.9	860	21	AAV13632	C. trachomatis pmp
43	844	17.7	866	21	AAV13638	C. trachomatis pmp
44	834.5	17.5	483	20	AAV34609	Chlamydia pneumoni
45	819	17.1	494	20	AAV34615	Chlamydia pneumoni

#### ALIGNMENTS

RESULT 1

AAW88421 standard; Protein; 928 AA.

AAW88421:

26-APR-1999 (first entry)

Chlamydia pneumoniae surface exposed protein Omp8.

Omp8: outer membrane protein 8; surface exposed protein; antigen;

Infection; diagnosis; vaccine; atherosclerosis; asthma.

Chlamydia pneumoniae.

MO9858953-AA2.

30-DEC-1998.

19-JUN-1998; 98WO-DK00266.

23-JUN-1997; 97DK-0000744.

(BIRK/) BIRKELUND S.

(CHR1/) CHRISTIANSEN G.

Birkelund S, Christiansen G, Knudsen K, Madsen A;

Mygind P;

WPI: 1999-105610/09.

N-PSDB; AAX06820.

Species-specific test for identifying mammals infected with

Chlamydia pneumoniae - comprises detecting antibodies specific for

[illegible]

RESULT 15  
E72130  
polymorphic membrane protein G family CP0761 [imported] - Chlamydophila pneumoniae (strain  
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: E72130; G81541  
R:Kalmun, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.  
Nature Genet. 21 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A/2000; MUID:99206606  
A:Accession: E72130  
A:Molecule type: DNA  
A:Residues: 1-841 <ARN>  
A:Cross-references: GB:AE001586; GB:AE001363; NID:94376263; PIDN:RAD18172.1; PID:943762  
A:Experimental source: strain CWL029  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; McClary, J.; McClary, G.; Salzberg  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis Morn and Chlamydia pneumoniae AR39.  
A:Reference number: AB1500; MUID:20150255  
A:Accession: G81541  
A:Molecule type: DNA  
A:Residues: 1-841 <REA>  
A:Cross-references: GB:AE002233; GB:AE002161; NID:97189672; PIDN:AAF38561.1; PID:971896  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: pmp\_2; CP0761

Query Match	32.7%	Score 1562	DB 2	Length 841
Best Local Similarity	39.2%	Pred. No. 2.5e-77		
Matches 369	Conservative 121	Mismatches 312	Indels 140	Gaps 20

[illegible]

Db	179	-----	AGSLTIANKSLSF	193
Qy	249	VTGASSSTVTDGMSGCAICAYCTSTDTVVTLTGNOMLFSSNNTSTTGAAGIYVKLELASC	308	
Db	194	-----	-----GNSSTRCGAIHTKNTLTSSG	214
Qy	309	GLTLEFSRNVGNGAP-----KGALAIEDSGELSLADSDGIYVLTGVTSTPGCT-NRS	363	
Db	215	GETLF-----OGNAPPAAGKGGAIALADSGTSLISDSDGIIIEGNTIGCT-GIVSHS	267	
Qy	364	SIDLTGSAKMTALSAAGRAIYFDPDPTTSSSTVTVDLVKVNETPADSALOYTGNILFTG	423	
Db	268	AIDGTSAKITALLAAGQHTIYFDPDPTVWGTSVADALNINSPDGTDNKEGTYGVFSG	327	
Qy	424	EKLESTEAADSKNLTSLKLDPYTLTSGTSLSKHBYTLQTOAQFOQADSKLEMDYGTLLP	483	
Db	328	EKLFEAEKDEKKNFTSKLDLVANAFKNTVYALKGVVLSANGFSDOANSKILMDGTSL-V	386	
Qy	484	ADTSTI--NNEVNISSIDGAKKAKIETKATSKNLTLSGVTTLTLDPTGTFVENHSLNPQ	541	
Db	387	ANTSIELTNETIIMIDSLRNKKIKLSAARQKDIRIDRPVYVALIDESTYQCNFLMEDH	446	
Qy	542	SYD-ILELKASGTVTSPAATVPDPIMGEKFKHYGQGTGWPJWGSTGASTATFNWTKTYI	600	
Db	447	SYDILLETLDAGKDVIYASDSRS-IDAVQSPRYGQGM-TINMSTD-DKKAATVSAKOSFN	503	
Qy	601	PNPRTIGSIYVNSLMMAFIDISSLIHMETRANESLOGDRAPWCAGISFNPHKDSKTRRG	660	
Db	504	PTAQBQAPLVYNNLMGSTRIDVRSQNFTELGTBGAPEYKRRFWAGISNVLIHRSRENRQK	563	
Qy	661	FRHLSSGGVYIGGNLHTSSDKTILSAFQOLFGRDIDYFVAKNOGTIVYGTLTYOHNETYIS	720	
Db	564	FRHYSGGVAVGASIRMPGCDPLTSLGFAQLFARDKDYDMNNTNFAKTAYGSLRLQHDASLVS	623	
Qy	721	LPC-----KLKPCSLSYVPELEIYVLESGNLSHTHTNDLTKY-----TTYPRVAGSWG	769	
Db	624	VVSILTBEGGIRELLIIPYYSKTLPCSYGOLSYGHTHRMKRTSLPBPPTLSTDDHTSWG	683	
Qy	770	NDSFALFERGGRAPICLDBESA---LFEQYPMFKLOFYAHQOEGKPDGTFAREFGSSRLV	826	
Db	684	GYYVAGELGTR--VAVENTSGRGFOERYTPPVAKQAYAAQADSDVELGALSTRDPSHLV	741	
Qy	827	NLALPIGIRFEDKESDODATVNLTLGYTVDLVRSNPDCITTLTIRISGDSWKTFGTNLARQ	886	
Db	742	NLALPIGLIKLEKR--FAYOYHYVAAVMSPDVCRSNPCKCTTLLSNQGSWKTKGSNLARQ	799	
Qy	887	LVLRAGNHFCNSNFEAFSOFSEFLTRSSNRNYNDVLAKTOF 928		
Db	800	GIYVASGFRSLGAAALFPGNNGFEMRRSSSYNVADAGSKTF 841		

Search completed: October 2, 2001, 03:29:20  
Job time: 1337 sec

R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.  
A:Reference number: A72000; MUID:99206606  
A:Accession: D72077  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-928 <ARN>  
A:Cross-references: GB:AE001628; GB:AE001363; NID:94376730; PID:AAID8593.1; PID:9437673  
A:Experimental source: strain CML029  
C:Genetics:  
A:Gene: pmp\_11

Query Match 38.88; Score 1853; DB 2; Length 928;  
Best Local Similarity 42.68; Pred. No. 4.5e-93;  
Matches 404; Conservative 171; Mismatches 332; Indels 42; Gaps 20;  
1 MKSFPKVFSTFAIFP--LSMIATETVLDSASPDGN-KGNFVSRESQDAGTTLYFK 57  
1 MKTSPWLVSVLAFSCHLOSILANEELSPDSEFNGINDSGTTPKTS---ATTYSLT 56  
58 GNVLTENIPGCTAATKSCFNNTKGLDFTGNGNSLRFQYDAGTAVAGAVSSVVDKST 117  
57 GDVEFFEY- PGKGTPLSDSCFKQTDNLTFLGNGHSLRFPGFDAGHAGAA- STTANKNL 114  
118 TFIQFSSLSFIASPGSSITTTGKAVSCSTGSLKFDKNVSLFKNFSTDNCGAITAKTLS 177  
115 TFSGFSLSLSPSSSTVYTTGOGTLS-SAGVNLLENIRKLVAGNFSTADGAIKASFL 173  
178 LTGTTMSALFSENTSSKKGAIGTSDALTITGNOGEVFSFSDNTSSDSGAIFTEASVTS 237  
174 LTGISGALFENSSSTYKGAIAATTAGARIANNNGYVFLNIASTSGAIDDEGTSLIS 233  
238 NNAKVSFDNKVYTGASSSTGDMSGAICAKYTDRTKVTLTGNOMLLFSNNTSTTAGA 297  
234 NNKFLYF-----EGNAAKT-----GGAICNTKASGSPELIISNNKTLIFASVNAETSGA 284  
298 IYVKLELASGGLLFSRNSVNGTAPKGAIAIEDSGELSLSDSGDIVFLGNTVTS- 356  
285 IHAKKALSSGGEFFELNNVSSAT-PKGAISIDASELSLSAETGNTITVRNLTITTG 343  
357 -TPGTNRSSIDLTSAKKTALRSAGRAIYFDPITGSSSTVYTDVLEKVNTPADSALOY 415  
344 STDPPKRAINIGSNGKTELRANKNHTIFFYDPT--SEGTSSDVLKINNGSAGALMPY 401  
416 TGNITFGEKLETPADSKNLTSKLLQPTVLSGGTSLKAGVTLQTOAFTQOADSRLM 475  
402 OGTLIFSGEFTLADLEIKVADMKSFTQPVLSGGKLLQKGVLTLESTFSQEGSLTGM 461  
476 DVGTTLE-PADTSTNNVYINMISSIDGAKKAKIEKATSKNLTLSGTTLLDPTGTVEY 534  
462 DSGTTLSTAGSITTTNGINVDLSGLKQPVSLTRAKGASNKVIYSGKLNLDIDEBNITES 521  
535 HSLNPOSYDILELKASGTVTS---TAVTPDPIMGEKFFHYGOGTGWPIYWGTCASAT- 589  
522 HMEBHDQFLSKLTIVADVDYTNDISLIPVAPADPSEKGFQOGQNN-VMMTDTTANT 580  
590 --AFTNMTKTYIPNPERIGSLVPSNLMNADIDISSLHYMETANEGIQDRAFCAGLS 647  
581 KEATATWTKTGTFVSPERKSALVNTCMTLNGVFTDIRSLQLOLEIGATGMEHROGFVSSMT 640  
648 NEFKDKTKTRGRHRHLSGCVIVIGNLTCSDKILSAFQCLFGDRDYPFAKNOGIYTG 707  
641 NEFKHTGDENKRRHSGGVIGGSAITPPKDDLTFFAFCHLFAADKCCFLAHNNSRTYG 700  
708 GTLYYOHNET-----YISL-PCKLRPCSLSVYPTIEIPVLFNGNSYTPHTDNDLTKRYTY 761  
701 GTLFFKSHTLQOPNYLRLGAKFSESATIEKFPREIPALDVOVSFSHSDRMETHHTSL 760  
762 PTYKSGWGNDSFALEFGRAPICL-DESALFEQYMPFKALOYVAHOGEGFEGGTAEAREF 820  
761 PESGSMNSNECTIAGIGIDLPFVLSNPHLKFTEFLPKVKVMYVVSQNSFFESSDGGKF 820

821 GSSRLVNTALPIGIREDESCQDA-TYNLTGTYVDLVRSNPQCTTLRLISGDSMKTFG 879  
821 SIGRLNLTSPVGAKE-VQGIDGISTYDLSGFVSDYRNNPSTALVAVSPSMKIRG 879  
880 TNLARQALVLRAGNHFCFNSNFEAPSCFELRGSSRRYNDLCAKYOF 928  
880 GNLSTQALRLGNSNNYVNSNCELFHGYAMELRGSSRRYNDVYKILRF 928  
RESULT 14  
F81591  
polymorphic membrane protein G family CP0302 [imported] - *Chlamydia pneumoniae* (s  
C:Species: *Chlamydia pneumoniae*, *Chlamydia pneumoniae*  
C:Date: 31-Mar-2000 #sequence-revision 31-Mar-2000 #text-change 11-May-2000  
C:Accession: F81591  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39  
A:Reference number: AB1500; MUID:20150255  
A:Accession: F81591  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-949 <REA>  
A:Cross-references: GB:AE002192; GB:AE002161; NID:97189226; PID:AAF38159.1; PID:9718  
A:Experimental source: strain AR39, HL cells  
A:Genetics:  
A:Gene: CP0302  
Query Match 38.88; Score 1853; DB 2; Length 949;  
Best Local Similarity 42.68; Pred. No. 4.7e-93;  
Matches 404; Conservative 171; Mismatches 332; Indels 42; Gaps 20;  
1 MKSFPKVFSTFAIFP--LSMIATETVLDSASPDGN-KGNFVSRESQDAGTTLYFK 57  
22 MKTSPWLVSVLAFSCHLOSILANEELSPDSEFNGINDSGTTPKTS---ATTYSLT 77  
58 GNVLTENIPGCTAATKSCFNNTKGLDFTGNGNSLRFQYDAGTAVAGAVSSVVDKST 117  
78 GDVEFFEY- PGKGTPLSDSCFKQTDNLTFLGNGHSLRFPGFDAGHAGAA- STTANKNL 135  
118 TFIQFSSLSFIASPGSSITTTGKAVSCSTGSLKFDKNVSLFKNFSTDNCGAITAKTLS 177  
115 TFSGFSLSLSPSSSTVYTTGOGTLS-SAGVNLLENIRKLVAGNFSTADGAIKASFL 194  
136 TFSGFSLSLSPSSSTVYTTGOGTLS-SAGVNLLENIRKLVAGNFSTADGAIKASFL 194  
178 LTGTTMSALFSENTSSKKGAIGTSDALTITGNOGEVFSFSDNTSSDSGAIFTEASVTS 237  
195 LTGISGALFENSSSTYKGAIAATTAGARIANNNGYVFLNIASTSGAIDDEGTSLIS 254  
238 NNAKVSFDNKVYTGASSSTGDMSGAICAKYTDRTKVTLTGNOMLLFSNNTSTTAGA 297  
234 NNKFLYF-----EGNAAKT-----GGAICNTKASGSPELIISNNKTLIFASVNAETSGA 305  
298 IYVKLELASGGLLFSRNSVNGTAPKGAIAIEDSGELSLSDSGDIVFLGNTVTS- 356  
306 IHAKKALSSGGEFFELNNVSSAT-PKGAISIDASELSLSAETGNTITVRNLTITTG 364  
357 -TPGTNRSSIDLTSAKKTALRSAGRAIYFDPITGSSSTVYTDVLEKVNTPADSALOY 415  
365 STDPPKRAINIGSNGKTELRANKNHTIFFYDPT--SEGTSSDVLKINNGSAGALMPY 422  
416 TGNITFGEKLETPADSKNLTSKLLQPTVLSGGTSLKAGVTLQTOAFTQOADSRLM 475  
423 OGTLIFSGEFTLADLEIKVADMKSFTQPVLSGGKLLQKGVLTLESTFSQEGSLTGM 482  
476 DVGTTLE-PADTSTNNVYINMISSIDGAKKAKIEKATSKNLTLSGTTLLDPTGTVEY 534  
483 DSGTTLSTAGSITTTNGINVDLSGLKQPVSLTRAKGASNKVIYSGKLNLDIDEBNITES 542  
535 HSLNPOSYDILELKASGTVTS---TAVTPDPIMGEKFFHYGOGTGWPIYWGTCASAT- 589

Db 178 KTLISGNTSITFTSNKAKLGALYSSAASISGNTGOLVFMNNKGETGCGALGFAS 237  
QY 234 VTIISNAKSPFDKNVTGSSSTGDMGCAICAKTSTDTVTILGNMFLPSNNTSTT 293  
Db 238 SSITQSSSIFEGNATPDAG-----KGAICYCEGTGETPTLTISGNSLTFASNESSVT 291  
QY 294 AGCAIVKKELEASGGLTFSRNSVNGTAPKGAIAIDSGELSLSDSGDIVELGNTV 353  
Db 292 QGGAICAHLDLSAGPTLFSNNRCGNTAAGKGAIALDSSLSLSANQSGITFLGNTL 351  
QY 354 TSTT-PCNRSSIDLTGSAKMTALNSAGRAIFYDPI---TTGSTTVTVLKVNEIPA 409  
Db 352 TGTSAFTSTRNAYIGSSKATYNLRAAGOSIFYFDPIASNTTGA---DVLTIQOPDS 407  
QY 410 DSALOYGTIIFTEKLESTEAADSKNLTSLKLOPVTLSGGLSLKHGVTLOTOAFTQA 469  
Db 408 NSPLDTSIGVTEGKLSADEKADNFTSLKOPALASGLALKGNELVNGTQTE 467  
QY 470 DSRLEMDVTTLEPADTSTIN--NLVINISSIDAKKAKIETKATSKNLTLSGTTLLDP 527  
Db 468 GSTLMQPTKTK-ADTEAISTLKLIVDLSALEGNKSVSIETAGANKTTTLSPLVFQDS 526  
QY 528 TGTFFENHSLRNPQSTDLIELKAS-----GTVTSTAVTPDPIMGEKPHYGTQWIG 578  
Db 527 SCNFYESHITINQAFQTOPLVFTAAASDIYDALITSPVQTPPEP-----HYGYQHWE 580  
QY 579 PIYWGASTTAPFNTKTGYIPNPERIGSLVPSLSMNAFIDISSLHYLMEANEGLOD 638  
Db 581 ATWADTSTKSGSTMTWTTGYNPERRASVVDLSMASTFIRLQOQMTSOANSITQO 640  
QY 639 RAFWCAISNPFHKDSTKTRGCFRHLSCGYVIGNLHTSCDKILSAFCOLGRPDYEV 698  
Db 641 RGLMAGSTANFPHKXKSGNQAFRHKSXYIYGSAEDSENFSAFQOLRKGKDLFI 700  
QY 699 AKNQGVYGGTLYQHNETHYISLPCKLRPC--SLSTVPEIYVLSGNSLSTHTDNDLT 756  
Db 701 VENTSHNTYASLYLQHRALGGLP--MPSFGSITDMLDIPILNAOISYTKRNDMT 757  
QY 757 KTTYPTVSGMNSGFALFEGGRAPICL-DESALEFOYMPKLOFYVAHDEGFEQGT 815  
Db 758 RTTSTPEAGSTNNNGALELGGSLALYLPKRAPFPQGFPLKQOAYVSRQONRESA 817  
QY 816 EAREFGSSRLVNLALPIGIRFDESDCODATYNLTGYTVDLVRSNPDCTTLRLISGDSW 875  
Db 818 EARAFFDDGLVNCISYVIGIRLEKISDEKKNPEISLAYIGDYVRKPNRRTSLMWSGASW 877  
QY 876 KTFPGTUALQALVLRAGNHFCSNFEARSQSFSELRGSSRNINVDLAKYOP 928  
Db 878 TSLCKNLARQAFLASGSHLTLSPHVELSGEAYELRGSANHYNVDCGLRYSF 930  
RESULT 12  
H86546  
polymorphic outer membrane protein G family [imported] - Chlamydia pneumoniae (strain  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: H86546  
R:Shirai, M., Hirakawa, H., Kimoto, M., Taduchi, M., Kishi, F., Ouchi, K., Shiba, T., Is  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae j138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: H86546  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-928 <STO>  
A:Cross-References: GB:BA000008; NID:g8978822; PIDN:BA098658.1; GSPDE:GN00142  
A:Experimental source: strain j138  
A:Genetics:  
A:Gene: omp\_11

Query Match 38.8%; Score 1853; DB 2; Length 928;  
Best Local Similarity 42.6%; Pred. No. 4,5e-93;

Matches 404; Conservative 171; Mismatches 332; Indels 42; Gaps 20;  
QY 1 MKSSPFRVVFSPALFP--LSMIATEVYLDSSAFDGN-KNCFNSVRSQEDAGTYLFK 57  
Db 1 MTSIPWLVVSVLAFSCHLOSLANEELSPDSFNGNDSTFPKTS---ATTISLT 56  
QY 58 GNVLTENIPGTATATKSCFNNTKCDLFTTGNGNSLLFQTVAGTVAGAAVSVVDKST 117  
Db 57 GDVFFYE-PGKOTPLSDSCFQKOTTNLFTLNGHSLTFEGIDAGHAAGAA-STANKNL 114  
QY 118 TTFGSSLSFASPOSSITTGKAVSCSTGSLKFDKNNSLFSKAFSTDNGCAITAKTIS 177  
Db 115 TFCGSLTFSDSPSTVTVTGGTSL-SAGGYNLNIRKLVYAGNFTADGGALIGASPL 173  
QY 178 LTGTMSALFSENTSSKKGAIOTSDALTITGNOGEVSPSDTSSDSGAITFEASVTS 237  
Db 174 LTGTSGDALFSSNNSSYTGGAIAITAGARIANNGTGYRFLSIATSGALIDEGSTIS 233  
QY 238 NNAKVSFIDNKVYGASSSTTGDMSGAICAYKTSYDTKVTLTGNOMLFSSNNTSTTAGA 297  
Db 234 NKKFLYF-----EGNAKKT---GALCNTRKASGSPELIISNNKTLIFASVVAETSGA 284  
QY 298 IYVKLEIASGGLTFSRNSVNGTAPKGAIAIDSGELSLSDSGDIVELGNTVST- 356  
Db 285 IHAKRLAASGGTFEFLRNWVSAT-PKGAISIDASGELSLSAETGNTTFVRNTLTG 343  
QY 357 -TPGNRSSIDLTGSAKMTALNSAGRAIFYDPIPTTGSTTVTVLKVNEIPASALOY 415  
Db 344 STDYPRRNAINIGSKGFELPAANHTEFFYDPI--SEGTSSDVLKKNNGSAGALNDY 401  
QY 416 TGNIIPTGKLESETPAADSUNKLTSKLOPVTLSGGLSLKHGVTLOTOAFTQOASRLM 475  
Db 402 QGTILFSGETLADDELKAYADNKSFPQVSLSGKILLQKVLTLESFSGEAGSLG 461  
QY 476 DVGITLLE-PADSTINNLVYNISIDGAKKAKIETKATSKNLTLSGTTLLDPGTFTN 534  
Db 462 DSGTLSTTAGSITTYNLGINDSLGQPSVLTAKGASKVYAGKGLLIDIEINIES 521  
QY 535 HSLRNPQSTDLIELASGVTS---TAVTPDPIMGEKPHYGYGTWGPYVGTGASTT- 589  
Db 522 HMFSDQLFSLKTIYVDADVDTNVDISLIPYPAEDPNSEYFOGOMN-VNMTTITATMT 580  
QY 590 --ATPNMTKTGYIPNPERIGSLVPSLSMNAFIDISSLHYLMEANEGLOGRAFCAGIS 647  
Db 581 KATATMTKGTGVSPSPERKSALVCMVTFDIRSLQOLVEIGATGEMHKGFVWSMT 640  
QY 648 NFPHDSTKTRGCFRHLSCGYIYGNLHTCSDKILSAFCOLGRDXYFVAKNGTYVG 707  
Db 641 NFLHKTGDENRKGFRHTSGYVIGSAHTPKDDLTFEAFCHLFARDKCFIAHNSRTYG 700  
QY 708 GTLYYQHNET-----YISL-PCKLRPCSLYVTEIPVLFSGNLSYTHDNDLTKRTTY 761  
Db 701 GTLFFRHSHTLOPNTYLGRKAKFSBSATEKPREIPLALDVQVFSHSDNMETHYISL 760  
QY 762 PTVKSGWGNDSFALFEGGRAPICL-DESALEFOYMPFKLOFYVAHDEGFEQTEAREF 820  
Db 761 PESBSWMSNECIAGGIGDLDFVLSNPHLFTFIPQKMEVAVYVSQNSFESSDGRGF 820  
QY 821 GSSRLVNLALPIGIRFDESDCODA-TYNLTGYTVDLVRSNPDCTTLRLISGDSKFTG 879  
Db 821 STGRNLNLSIPYKAF-VQGDIGDSTYDLSQFVSDYRRNPQSTATLWSPDSMKING 879  
QY 880 TNLARQALVLRAGNHFCSNFEARSQSFSELRGSSRNINVDLAKYOP 928  
Db 880 GULSKQAFILRGSNNYVYNKCELFGHYAMELRGSSRNINVDVGIKLR 928

RESULT 13  
D72077  
polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: D72077



A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AAF38165.1; PID:g718923  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: CP0308

Query Match 40.8%; Score 1950; DB 2; Length 936;  
Best Local Similarity 43.1%; Pred. No. 2.5e-98;  
Matches 408; Conservative 179; Mismatches 330; Indels 30; Gaps 16;

1 MKSSFKFVPSFAIR-PLSMATEVYDSS-ASFDGNKNGFVSRESOE-DAGTYLER 57  
1 MKSSVWMLFEFSSIPLEFSSISIVAAEVLDSNNNSYDGSNGTFTVSTDDAAGTYSLL 60  
58 GNVLTENIPGTATIRKSCFNNKGDLEFTGNGNSLLFQYDAGTVAGAAVNSVYDKS 117  
61 SDVSFQNALGELPLASGCELEAGDLPFGNQHAKLFAFINNGSSAGVASTADKML 120  
118 TFGFSSLSFASPCSSIT-TGKAVSCSTGSLKFPKNSILFSKNEFTDNGAIFAKTL 176  
121 LFMDFSRILSTISCPILLSTPTGCAK-SVGNLSLTGNSQIIFTOQFSSDNGVITKNF 179  
177 SLTGTTMSALFESNT--SSKKGAIOTSDALITGNOGEVSESDNTSSDGAIFTEASY 234  
180 LLSGTQSFASFSRNOAFQKOGGVYATGTITENSPGIATFESQNLAKSGSGALYSTDNC 239  
235 TISNNKAVSFIDNKVYAGSSSTTGDMSGAICAKYKSTPTKVTLPFGNMLFENNSTTA 294  
240 SLTDNQVIFDGNASAEAOA-----QGALIC--TTTDVYTLTKNKLSTNNATLY 292  
295 GGAIVYKLELASGGLTFSRNSVNGTAPK--GGAIAIEDSGELSLASDGDVFLGNT 352  
293 GGAISGLKVISAGCPPLFOSN-IGSSSAGGGGAINIASAGELASATSGDITENNNO 351  
353 VNSTGTNRSSIDLTGSAKMTALBSAARATFYDPITGSSSTYTVDLKAVETPADA 412  
352 VNGSTST-RNALNIDITAKVISIRATGOSIYFDPITNPGTAOSTDLNLADANSE 410  
413 LQYTGNIIFTEKLESTEADSKNLSKLTQPYLSGGLSLKHGVTLQTOAFTQOASR 472  
411 IERGAIVFSGEKLSPTFERAIANVSTIRQPAVLARGLVLRDGYTVFVKDLTQSPGR 470  
473 LEMDGTTLLEPADTS-TINNVLINISSIDGAKAKIETKATSKNLTSTTLLDPTGTF 531  
471 ILMDCGTTLSAKEANSLNGLAIVNLSDGTNKAALTEADKNISLSGIALIDTEGSP 530  
532 YENHSIRNPOSYIILK---ASGYTSTAIVTPDPIMGKEFHGYGTGMPYWGTCAS 588  
531 YENHNLSKASTYPLELLETTAGANGTTLGALSTLQEPETHYGQGNM-QLSWANATSS 589  
589 -TATFMMWTGTYIPNPERIGSLVPSNLMAVAFIDISSLHYLMEETANGLOGDRAFWCAGLS 647  
590 KIGSIMTTRGTYIPSPERKSNLPLNSLMGNFIDIRSIINOLIEKSSGEPERELMLSGTA 649  
648 NEFHKSDTYTRRGRHLSSGYVIGNLHTCSDKILISAFCQLEGRDRDYFVAKNOGTYVG 707  
650 NEFYRDSMPTRHGFRAHISGAYALGATATPAEDQLFAFCQLEPARRNMHTKNGHDYTG 709  
708 GTYXOHNETHYISLPCKL-----RCSLSYVTELEPVLFSGMLSTHNDNLTXTYTY 761  
710 ASLYFHTHTGLEDIANFLWKGATRAVWLSLSQIILPLSFDAKFSYLLHNDNHTKTYTNN 769  
762 PYVKGSMGNDSEFLERGRAPICLDESALFEQYMPMKLOFYVAHOEGKEQGTTEAREFG 821  
770 SILKSGMRNDAPFADGASLPFIASVPIYLKKEVEPVKQYITAHODDYERAEERAN 829  
822 SSRVLNALPIGIRDEKESDCCDAYNLTLGYTVLVRSPNDCSTTLTKISGDSWKTEGTN 881  
830 KSELINVEIPIGTFERDSSEKGYDTLTLMYLLDAYRRNPCKQTSILIASDANMAYGTN 889  
882 LARQALVLRAGNHFCSNFSEASOSFELRGSSRNNYNDLAKYOF 928  
890 LARQGSVRANHFQVNPHEIFGQFAFEVRSSRNNYNTNLGSKCF 936

RESULT 9  
D66546  
polymorphic outer membrane protein G family [imported] - Chlamydia pneumoniae (st  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: D66546  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: D66546  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-930 <STO>  
A:Cross-references: GB:BA000008; NID:g9978818; PIDN:BA98654.1; GSPDB:GN00142  
A:Experimental source: strain J138  
A:Gene: pmp-8

Query Match 40.4%; Score 1929; DB 2; Length 930;  
Best Local Similarity 43.4%; Pred. No. 3.4e-97;  
Matches 414; Conservative 164; Mismatches 327; Indels 48; Gaps 17;

1 MKSSFKFVPSFAIR-PLSMATEVYDSS-ASFDGNKNGFVSRESOE-DAGTYLER 56  
1 MKTLPKHLISLTVPIILSLTATYGDADSLSPDSFDGAGSGSTFPKSTADANGTNYV 60  
57 KGNVLTENIPGTATIRKSCFNNKGDLEFTGNGNSLLFQYDAGTVAGAAVNSVYDKS 116  
61 SGNVYI-NDAGKGTALTGCCFETETGDLFTFGKGYFSFETVDAGSNAGAA-STRADKA 118  
117 TTFGSSLSFASPCSSITTGKAVSCSTGSLKFPKNSILFSKNEFTDNGAIFAKTL 173  
119 LFTFGSNLSFIAAPGTVAAGSKSTLS-SAGALNLTNDGTITLFSQVNSSEANNGALITTA 177  
174 KTLSTGTTMSALFESNTSSKKGAIOTSDALITGNOGEVSESDNTSSDGAIFTEAS 233  
178 KTLISGNTSSITFTNSAKKLGATISSAASISGNTQOLVPMNKKETGGALGFAS 237  
234 VTISSNAKVSFIDNKVYAGSSSTTGDMSGAICAYKSTPTKVTLPFGNMLFENNSTTT 293  
238 SITQSSLSFESGNTATDAG-----KGAICYCEKTEGEPITLISGNKSLFPAENSVT 291  
294 AGGAIYVKKLELASGGLTFSRNSVNGTAPKRGGAIAIEDSGELSLASDGDVFLGNTV 353  
292 OGGAICAHGLDLSAAGPTLESNNRCGNTAAGKGAIAIADSGSLISANOGDITFLGNTL 351  
354 TSTPT-PGTRRSSIDLTGSAKMTALBSAARATFYDPI--TTGSSSTYTVDLKAVETPA 409  
352 TSTSAFTSRNRAIYLDSSAKITRLRAAGOSIYFPIPLASNTTGAS---DVLTRNOPS 407  
410 DSALQYTGNIIFTEKLESTEADSKNLSKLTQPYLSGGLSLKHGVTLQTOAFTQOAS 469  
408 NSPLDYSGTIVFSGEKLSADEAKAADNFTSILKQPLASGLTALKGVELDVGCTQJE 467  
470 DSRLEMDVTTLEPATYSLIN--NLVINISSIDGAKAKIETKATSKNLTSTTLLDLP 527  
468 GSTLMLQPGTKLK-ADTEAISLTKLVVDSALEGNKSVSIEAGAKKTTTLLSPLEFDS 526  
528 TGTGFENHSIRNPOSYDIELKAS-----GVTYSTAIVTPDPIMGKEFHGYGTGNG 578  
527 SGNFYSHRTINQAFIOPLVYFTRAAASDIYDALLTSSVQYPER-----HYGYGHNH 580  
579 PIYWGTAFTATFNNTKTYIPNPERIGSLVPSNLMAVAFIDISSLHYLMEETANGLOGD 638  
581 ATWADSTAKSGMTWVTGYNPNPERRASVYDPSLMASTFDIRLQIINTSQANSIYQO 640  
639 RAFWAGLSNFRHKDSKTRRGRFRLHSGGYVIGNLHTCSDKILISAFCQLEGRDRDYV 698  
641 RGLMASGTANFPHKDKSGTNOAFRRKSYGYIVGSAEDESENIISVAFQLEFGKDKDLFI 700

[illegible]

Query Match	40.8%	Score 1950:	DB 2:	Length 936:
Best Local Similarity	43.1%	Pred. No. 2.5e+98:		
Conservative 408:		Mismatches 330:	Indels 30:	Gaps 16:
QY	1	MKSSPFKVFESFAIF-PLSMATETVDDSS-ASPDGKKNKNSFVRSQE-DAGTYIFAK	57	
		:	:	:
Db	1	MKSSVSWLFFSSIPFSSLSIVAAEVLDDSSNNSDGSNGCTTFVFTTDDAAGTYSLL	60	
		:	:	:
QY	58	GNVTLEIPGCGTATKTRCFNNTKGDLFTFGNGSNLLFQTVDAAGIVAAANSSVVDKST	117	
		:	:	:
Db	61	SDVSRQNGALGIPLPLASGCELEAGDGLFTFGNOHMLKAFINAGSSACTVASTAADNLT	120	
		:	:	:
		:	:	:

```

QY 118 TTTGSSLSFJASPPSSITT-TGKAVSCSTGSLKRDKNVSLLEFKNPESTDNGCAITATLT 176
      | | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 LENDSSRLSISCPBILLSPTGQCALK--SVGNLSLTGSSQIILFPONFSSDNGVINTNPF 179
      | | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 SLTGTMSALFSEENT--SSKRGGAQTSDALITITNOGEVEFSDTSSDGAALFTEASV 234
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 180 LLSGTQSOPAFSRNQAFCNGQGVYATGTITIEHSPGIVSPSOILAMGSGGALYSTDNC 239
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 235 TISNNAKVSFIDNKVITGASSSTTGDMSGCAICATYSTDTKVTLTGNOMLLFNSNTSTA 294
      : : : : : : : : : : : : : : : : | : | : | : | : | : | : | : | : | :
Db 240 SITDNFOVIFDONSAMEAAQA-----OGGALIC--TTTDDKVTYTLGNMNLFTNTVALTY 292
      : : : : : : : : : : : : : : : : | : | : | : | : | : | : | : | : | :
QY 295 GGAIVYKRIELASGGTLTFSRNSVNGTAPK--GGAALDEGSELTSADSGDIYLCNT 352
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 293 GCAISGLAKYISAGPTLPQSN-IGSSAGOGGAGATINIASGELASATSGDITFNNO 351
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 353 VTSPTGTRSSIDLGTSAKMTALRSAAGRAIFYDPITTTGSSPTTVTDLVKNETPADSA 412
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 352 YNGSGTST-RNAINIIDRAKYSIRAPATGOSITFDPITNPQTASTPTLNLINLADANSE 410
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 413 LQYTGNILFTGKLSLSETPAADSKNLTSKLQPTVLTSGTSLSKHVTLTQTOAFQOQADR 472
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 411 IEYGGAIFYSGEKLSPTKEALAAVNTSTIRPAPVALRDLVROGVTYTFDILQSPGSR 470
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 473 LEMDQVTLTEPBDTS-TINNLIYINISSIDGAKKAKIEFKATSKNLTSGTITLLDPDTF 531
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 471 ILMDSGTISAKEAHLNSLGLAVNLSSLDGTRKALKTEADKKNLSGTLALIDTBSF 530
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 532 YENHSLRNPQSYDILK--ASGTVTSTAVTPDPIMEKEFNHYGYGTGPIVMGTGAST 588
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 531 YENHMLKSASTPRLLELTITAGANGTITLGALESTLTLOEPENHYGQGNM-OLSWANATSS 589
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 589 -TATFNWTKTGITIPPERIGSLVPRSLTNNAPFLDISLLEYMETNEGLQGRAPFMGSL 647
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 590 KIGSINWRTGTIIPPERKSNLPLNSLWGNFLDKINSIQLETSSSGPPERREIMLSGIA 649
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 648 NFPHNDSTKTRGRGFNHLSSGYVYTGGLNLTGSDKILISAFCOLLGRDRPYEFAKNGQVYTG 707
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 650 NEFYDSMPTHRGPFHISGGVALGITATTPADBDQTFACQLPARDRNHITGKNHGDYTG 709
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 708 GTLTYQHNETTISLECKL-----RPSLSVPTPTPIPLFSGNLSYTTDNDLTKTKTTY 761
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 710 ASLYRHHEGLEFDIANFLMGKATRAPWVLSLSIQIILPSIFAKSYTLTDDHMKTYTTDN 769
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 762 PIVKESWGNDSFALPEFGAPICIDESALFPOYMPFMLOLVYAHQEGFROGEAREPFG 821
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 770 SIIKSSMRDAPCADLGASLPVIVPVLKKEVEFFVAVOYIYAHQDPFYRYABGRAFN 829
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 822 SSRILVNALPIRDKESDCODATYTNLTIGYVLDVLRNSDCTTTTLRISGDSMKTEGTN 881
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 830 KSELINVEIPIGVTERDSKSEKGYDLDLWAIIDAYRRNPKQOTSILASDANMMAYGTN 889
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 882 IARQALVLRAGNHCFPNSNFEAFSOFELGSSSNYVVDGATYQF 928
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 890 IARQEFVRAAHNFQVNPHEIFGQFAEVSSSSNNYTNLNGSFCF 936
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```

RESULT 8  
BB1591  
polymorphic membrane protein G family CP0308 [Imported] - Chlamydia pneumoniae (s  
C.Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C.Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C.Accession: BB1591  
K.Reed, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A.Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39  
A.Reference number: AB1500; MUID:20150255  
A.Accession: BB1591  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-936 <REA>



```

RESULT 5
B72077
polymorphic membrane protein g family CP0306 [imported] - Chlamydia pneumoniae (stra
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: B72077; B81592
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: B72077
A:Molecule type: DNA
A:Residues: 1-928 <ARN>
A:Cross-references: GB:AE001628; GB:AE001363; NID:94376730; PIDN:AA018591.1; PID:9437673
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: B81592
A:Molecule type: DNA
A:Residues: 1-928 <REA>
A:Cross-references: GB:AE002192; GB:AE002161; NID:97189226; PIDN:AAE38163.1; PID:9718923
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: pmp_9; CP0306

```

```

Query Match 41.4%; Score 1977; DB 2; Length 928;
Best Local Similarity 45.1%; Pred. No. 8.3e-100;
Matches 428; Conservative 149; Mismatches 329; Indels 44; Gaps 19;

```

```

OY 1 MKSSPFKFEYSFAIFPLSM-----IATEVLDSSASFDGKNGNSVRESOEBA-GTT 53
    ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
    1 MKSSLHMFLLSSLLPLSLNFSAPAAVEINLGPNTSSG--PGHYTPPAQTNNADGTI 58
DB 54 YLEKGVNTLENIPTGTAATKSCFNNTKDDLTFTGNGNSLLQYVDAGVAGAAVNSVY 113
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    59 YNLGTGVSTN-AGSPFALTASCFTKTTGNLSFGHGQYFLLQNTDAG--AACTPTNTAA 115
OY 114 DKSTPFISSSLFASPGSSITTKGAVSCSTGSLKPKKNVSLFSKNEFTDNGAATA 173
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    116 NKLLSTSGSYSLIOT--TNATGTGAIR-STGACSIQSNYSCRYGQNFNDNGALOG 172
DB 174 KTLSTLGTMSALFSENTSSKKGALQTSDALTTGNGEVSFSDNTSSDSCAIFTTAS 233
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    173 SSISLS-LMNPILFAKNKATOKGALYSTGGITNTNLSASFSEMTAANNNGAITYTAS 231
OY 234 VITSNNAKVSFIDNKYTGASSSTTGMSGALCAVTSIDTKY-TLTGNOMLLFSNNTST 292
    ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
    232 SFTSSKKAISFTINNSTATA--TGGAICYSSSTSAKPKVLTLSIDNGELNFTGNVTAI 285
DB 293 TFGAGATVYKLELASGGLTFLSRNSVNGCTAPKGAIAIEDSGELSLSDSGDIYFLGNT 352
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    286 TSGCATYTDNLVLSGCPFLFKNNSAIDTRAAPLGAIALADSGSLSLAGDITTEGNT 345
OY 353 V---TSTPTGNTSSSIDG--TSAKMTALRSAGRAIYFPDPTTGGSSSTVTDLVAKVNEPT 408
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    346 VYKASSSGOTTNRNSININNTNAKIYQLRASOGNTIYFPDPTTSTTALADALNAGPD 405
OY 409 ADSALQYTNIIFTGKGLSETFAADSKNLTKLQPVTLTSGGLSKLHKVTLQTOFAFQO 468
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    406 LAGNPAYQGTITVFSGEKLEAEADNLSKSTIQDPLTLAAGQLSKSGVTLVAKSFSOS 465
DB 469 ADSRLMDVGTTLPEADSTINNLVINISIDGAKKAKIETKATSKNLTLSGTTILDDT 528
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    466 PGSTLMDAGTLETADGTTINNLVINLVSQIKETKATIKAQASQTVTLSSGLSLVDPS 525
OY 529 GFYEHSLRNPOSYDILEKAS--GTVTSTAVTTPDIPNGEKFYHYQGTWGPVWGTGA 586
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    526 GNYEEDVSNMNPQVFSCLTLTADDPANIHITDLAADPLKKNPIMHGYQGNMA-TSMQEDT 584

```

```

OY 587 ST---TATENMTKGYIPNPERIGSLVPSNLNATFIDISLHMETANEGLQGRATWC 643
    : | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    DB 585 AKKSAATLTWTKIGYNPERKGTLVANTLWGSFYVDARSIQDLVATKVRQSGEFTGWC 644
OY 644 AGLSNFPHKDSKTRGRFNRHLSGGVYIGNLTCSDKLTLSAFCOLFGDRDPYVAKNOG 703
    ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
    DB 645 EISNFPKHDSKTRKINNGFRHISAGYVAGTTTLASDNLNLTAFQCLFGDRDHFINKNA 704
OY 704 TYVGGTLYOHNERYISLPCKLKPCSLVTP--TEIPVLESGNLSYTHTDNDLTKTYTQ 760
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    DB 705 SAYAASLHLOHATLTS-----PSLRLTPGSESEQPLVFAQISYISKMTMTYYTQ 758
OY 761 YPTVAGSNGDSFALFEGGRAP-TCIDESALPEQYMPFKLQFYVAHOGEFKEGTE-AR 818
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    DB 759 APKGSWMYNDSCALFLASLPHLTALSHGLTHAFPFIKVASTYHDSFKERNITLVR 818
OY 819 EFGSSRLVNLALPIGRFEDKESDQDATYNTLGYTVDLVRSNPDCYTLRLISGDSWKE 878
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    DB 819 SFDSGDLINVSYPIGITGTFRFRSNERASYEATYIYADVIRKRPDCTYALLINTSMKTT 878
OY 879 GTNLARQALVLRAGNHEFCNSNFEAFSOFSELRGSSRYNVDIAKQYF 928
    ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
    DB 879 GYNLRQAGIGRAGLTFYAFSPNLEVTSLNLSMEIRGSSRSRYNADLGKQFQF 928

```

# RESULT 6

```

polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: C72078
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: C72078
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-936 <ARN>
A:Cross-references: GB:AE001627; GB:AE001363; NID:94376721; PIDN:AA018589.1; PID:9437
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: pmp_7

```

```

Query Match 40.8%; Score 1951; DB 2; Length 936;
Best Local Similarity 43.1%; Pred. No. 2.2e-98;
Matches 408; Conservative 179; Mismatches 330; Indels 30; Gaps 16;

```

```

OY 1 MKSSPFKFEYSFAIF-PLSMIATEVLDSS-ASFGKNGKNGNSVRESOE-DAGTYLYRK 57
    ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
    DB 1 MKSSVSWLFFSSLPFLSSLSIYAAEYTLDSNNSYDSNGTFTTYSTIDAAAGTYSLL 60
OY 58 GAVTLENIPTGTAATKSCFNNTKDDLTFTGNGNSLLFQTVDAAGVAGAAVNSVYDKST 117
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    61 SDVSRQACALGIPLASGCFLEAGGDLTFQGNQHALKFAFINAGSSAGVASTSADKNL 120
OY 118 TFISSSLFASPGSSIT-TGKAVSCSTGSLKPKKNVSLFSKNEFTDNGAITYAKTL 176
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    121 LFNDSRLIISCPILLSPTGQCALK-SVGNLSLNGNSQIIFTQNFSSDNGCVINTKPF 179
OY 177 SLTGTMTSALFSENT--SSKKGALQTSALTLTNGEVSFSDNTSSSGAAIFTEAVY 234
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    DB 180 LLSGTQSPAFSRNQAFTRKQGGVYATGTITENSFGVLSQNLAKSSGALSTDMC 239
OY 235 TISNNAKVSFIDNKYTGASSSTTGMSGALCAVKTSTDTKYTLGOMLLFSNNTSTA 294
    | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
    240 SITDNOQVIFDNGSAAEAAQA---QGALICC--TTTQKTYTLTKNKNLSFTNTALTY 292
OY 295 GCAIYKLELASGGLTFLSRNSVNGCTAPK--GGAIALIEDSGELSLSDSGDIYFLGNT 352
    ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : : ||| | : : : :
    DB 293 GGLISGLKVISISAGFPFLPQSN-ISGSSAGOGGGAINIASGELALSATSGDITENNNO 351

```



577 MGPIWGTGASTTATFNTKGTGYPNPERIGSLVPSNLMNAFIDISLHLYMETANESIQ 636  
421 WPIIWTGATSTAFNMTKGTGYPNPERIGSLVPSNLMNAFIDISLHLYMETANESIQ 480  
637 GDRAFKAGLSNFFKDKSTKTRGRFHLSSGYYVIGGNLHSCDKIISAFCOLLGDRDY 696  
481 GDRAFKAGLSNFFKDKSTKTRGRFHLSSGYYVIGGNLHSCDKIISAFCOLLGDRDY 540  
697 FVAKNGYGGGLYYOHNETYISLPCRLRPSLSVPEIYVLSGNSLYHTDNDLKT 756  
541 FVAKNGYGGGLYYOHNETYISLPCRLRPSLSVPEIYVLSGNSLYHTDNDLKT 600  
757 KYTTPYVKGWGNDSFALFEGGRAPICLDESALFEQYMPFKLOFYVAHOGEFKGTE 816  
601 KYTTPYVKGWGNDSFALFEGGRAPICLDESALFEQYMPFKLOFYVAHOGEFKGTE 660  
817 AREFGSSRLVNLALPIGIFPKDESCQDQATYNTLGYTYDLYRSNPDCTTLIRISDSK 876  
661 AREFGSSRLVNLALPIGIFPKDESCQDQATYNTLGYTYDLYRSNPDCTTLIRISDSK 720  
877 TGTGTLARQALVLRAGNHCFNSNFEAFSOFSELRGSSRNRYNDLGAKEYF 928  
721 TGTGTLARQALVLRAGNHCFNSNFEAFSOFSELRGSSRNRYNDLGAKEYF 772

RESULT 2  
G86546  
polymorphic outer membrane protein G family [imported] - Chlamydia pneumoniae (strain  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: G86546  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tsubuchi, M.; Kishih, F.; Ouchi, K.; Shida, T.; Is  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: G86546  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-928 <STO>  
A:Cross-references: GB:BA000008; NID:g8978821; PIDN:BA098657.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: pmp\_10

Query Match 43.3%; Score 2069; DB 2; Length 928;  
Best local similarity 46.9%; Pred. No. 8.4e-105;  
Matches 446; Conservative 141; Mismatches 319; Indels 44; Gaps 20;  
1 MMSPEKPEVF-STFAIF---PLSMATEVVLDSASFQGNKN-GNFSVRESQEDACTYTL 55  
1 MMSQFEMVLISLTLACFTSCVFAATAENIGPDSFDSSTNGTYTPKNT--TTGIDY 58  
56 FPGNVTLENIPTGTATITSCFNNTKGLITFTGNGNSLLFQYVDACTVAGAAVNSVVDK 115  
59 LGGDITLQNL-GDSALITKGFSDITESTLSFAKGYSLFELNKS-SABGAL-SYTTDK 115  
116 STFTFESSLSFASPGSSIT--GKAVSCSTGSLKPKNLSLSKPNSTNGAATA 173  
116 NLSITGFSSLTLAASSVITTPSGKAVKCG-GDLTFDNGITILEKODCEENGAIST 174  
174 KTLISLGTWMSALFSENTS---KKGAIQTSALTLITNGOGEVFSVDNTSDSGAIFT 230  
175 KNLSTKNGTSGISFEGNKSATGKGAICATGTYDITNNAPTLEPSNNAEAGAINS 234  
231 EASVITSNNAKYSFIDNKATYGASSSTTGDMSGAICATYSTDYKTYLNGNMLFSNNT 290  
235 TGCCTTGTGTSLSFESNSVT---ATAG---NGGAL-----SGDADVTISGNSVTSSEQ 283  
291 STTAGGAIYVKKLELAS--GGLTFFSRNSYNGGTAPKGAIAIEDSGELSLSDSDIYF 348  
284 AVANGAITYAKKLTLASGGGGISFSNNIVOGTTAGNGAISTILAEGCSLSAEDDITF 343

349 LGNTVSTTP-CTNRSSIDLGTSAKNTALRSAGRAIYVDPITGGSSVTYVDLVKNET 407  
344 NCAIATAYPQTTKRNISIDGSTAKITNLRASIGHSIFEYDPTANTADSTDTLNLKA 403  
408 PADSAIQYGNIIIFGEKLSFTEADSKNLQPTLSGGLSLKKGVTLLQQAFA 467  
404 DAGNSIDYSGSIVFSGEKLSEDEAKYADNLTSLKQPVTLTGNLVKAGVTLLKGFQ 463  
468 QADSRLEMOVGTLEBA-DSTIINNVLINISSIDGAKKAKIETKATSKNLITSGTYTLLD 526  
464 TNGSSVINDAGTTLKASTEEVTLTGLSIPVDSIGBEKKVYVIAASAKVVALSGPILLDD 523  
527 PGTVEYENSLRNPOSDILLETAKSGTSTAATPDPIMGEKFFHYGQGTWGPYV- 582  
524 NCGNAVENHDLGKTOPFSFVQSLGATTTDPAVPATPAPHYGQGTWGC-MTWAVDT 582  
583 -GTGASTTTFNMTKGTGYPNPERISLVPSNLMNAFIDISLHLYMETANESIQDRAF 641  
583 ASTPKRTATLTAMTNGYLPNPERQPLVNSLWGSFSDIOAIQVIERSLATLCSDRGF 642  
642 WCAGLSNFFKDKSTKTRGRFHLSSGYYVIGGNLHSCDKIISAFCOLLGDRDYVAKN 701  
643 WAAGVANFLDKDKKGRKRYRHSKGYAIGAAQCSNMLISFAFCOLLGDRDYVAKN 702  
702 OGTVYGGTLYYOHNETYISLPCRLRPSLSVPEIYVLSGNSLYHTDNDLKTXY 758  
703 HDTYVAGAYIYQH---ITGSGFICGLDKLPGSMKSHKPIVLEGLAVSHVSNDLKTY 758  
759 TTPYVYKSGWGNDSFALFEGGRAPICLDESALFEQYMPFKLOFYVAHOGEFKGTEAR 818  
759 TTPYVYKSGWGNDSFALFEGGRAPICLDESALFEQYMPFKLOFYVAHOGEFKGTEAR 818  
819 EGSSRLVNLALPIGIFPKDESCQDQATYNTLGYTYDLYRSNPDCTTLIRISDSKMTF 878  
819 SEDDSLWFLNLSPIGKFKFKFSDCNDYSDYLDLIRNDPKTTTALVISGASMEY 878  
879 GNTLARQALVLRAGNHCFNSNFEAFSOFSELRGSSRNRYNDLGAKEYF 928  
879 ANMLARQALVLRAGNHCFNSNFEAFSOFSELRGSSRNRYNDLGAKEYF 928

RESULT 3  
G81591  
polymorphic outer membrane protein G family CP0303 [imported] - Chlamydia pneumoniae (s  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: G81591  
R:Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salze  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MOpn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255  
A:Accession: G81591  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-928 <REA>  
A:Cross-references: GB:AE002192; GB:AE002161; NID:g7189226; PIDN:AAF38160.1; PID:g718  
C:Genetics:  
A:Gene: CP0303

Query Match 43.3%; Score 2069; DB 2; Length 928;  
Best local similarity 46.9%; Pred. No. 8.4e-105;  
Matches 446; Conservative 141; Mismatches 319; Indels 44; Gaps 20;  
1 MMSPEKPEVF-STFAIF---PLSMATEVVLDSASFQGNKN-GNFSVRESQEDACTYTL 55  
1 MMSQFEMVLISLTLACFTSCVFAATAENIGPDSFDSSTNGTYTPKNT--TTGIDY 58  
56 FPGNVTLENIPTGTATITSCFNNTKGLITFTGNGNSLLFQYVDACTVAGAAVNSVVDK 115  
59 LGGDITLQNL-GDSALITKGFSDITESTLSFAKGYSLFELNKS-SABGAL-SYTTDK 115

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2001, 03:29:09 ; Search time 76.79 Seconds  
(without alignments)  
920.562 Million cell updates/sec

Title: US-09-446-677b-10

Sequence: 1 MKSFEKFEVSTFAIFPLSM.....FELRGSSRMVNDLGAKEYQF 928

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3991	83.5	772	2	H86492 - Pmp_3 [Imported] -
2	2069	43.3	928	2	G86546 polymorphic outer
3	2069	43.3	928	2	G81591 polymorphic outer
4	1977	41.4	928	2	E86546 polymorphic outer
5	1977	41.4	928	2	B72077 polymorphic outer
6	1951	40.8	936	2	C72078 polymorphic outer
7	1950	40.8	936	2	C86546 polymorphic outer
8	1950	40.8	936	2	B81591 polymorphic outer
9	1929	40.4	930	2	D86546 polymorphic outer
10	1929	40.4	930	2	A81591 polymorphic outer
11	1925	40.3	930	2	D72078 polymorphic outer
12	1853	38.8	928	2	H86546 polymorphic outer
13	1853	38.8	928	2	D72077 polymorphic outer
14	1853	38.8	928	2	F81591 polymorphic outer
15	1562	32.7	841	2	E72130 polymorphic outer
16	1442	30.2	1276	2	B86546 polymorphic outer
17	1442	30.2	1276	2	C81591 polymorphic outer
18	1435.5	30.0	922	2	B72131 polymorphic outer
19	1435.5	30.0	922	2	E86491 polymorphic outer
20	1434.5	30.0	922	2	F81539 polymorphic outer
21	1381.5	28.9	1407	2	B72078 polymorphic outer
22	1377.5	28.8	973	2	B86547 polymorphic outer
23	1377.5	28.8	973	2	F72076 polymorphic outer
24	1377.5	28.8	995	2	C81593 polymorphic outer
25	1235	25.8	712	2	E86492 polymorphic outer
26	1115.5	23.3	1013	2	G71460 polymorphic outer
27	1053	22.0	967	2	H81722 polymorphic outer
28	1017.5	21.3	445	2	E86493 polymorphic outer
29	916	19.2	359	2	C86493 polymorphic outer

30	905	18.9	867	2	F81721 polymorphic outer
31	854	17.9	878	2	B71460 polymorphic outer
32	819	17.1	494	2	D86493 polymorphic outer
33	801	16.8	427	2	A86493 polymorphic outer
34	754	15.8	186	2	G86492 polymorphic outer
35	680.5	14.2	1609	2	A86611 polymorphic outer
36	680.5	14.2	1609	2	A86611 polymorphic outer
37	672.5	14.1	1978	2	H72013 polymorphic outer
38	670	14.0	947	2	G72076 polymorphic outer
39	667.5	14.0	978	2	B81593 polymorphic outer
40	667.5	14.0	978	2	B81593 polymorphic outer
41	666	13.9	947	2	D72067 polymorphic outer
42	664	13.9	1723	2	H86557 polymorphic outer
43	664	13.9	1723	2	E72067 polymorphic outer
44	664	13.9	1732	2	C81601 polymorphic outer
45	663.5	13.9	946	2	C86549 polymorphic outer

## ALIGNMENTS

## RESULT 1

H86492  
Pmp\_3 [Imported] - Chlamydia pneumoniae (strain J138)  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: H86492  
R:Shirai, M.; Hirakawa, H.; Kinoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: H86492  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-772 <STO>  
A:Cross-references: GB:BA000068; NID:98978389; PIDN:BAA98226.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: pmp\_3\_2

Query Match	83.5%	Score 3991	DB 2	Length 772
Best Local Similarity	99.9%	Pred. No. 2.5e-209		
Matches 771; Conservative 1; Mismatches 0; Indels 0; Gaps 0;				
QY 157	LLFSKNSTNDGGAITAKTSLTGTMSALFSENTSSKGAIGTSDALTTGNGEVSF	216		
DB 1	MLFSKNSTNDGGAITAKTSLTGTMSALFSENTSSKGAIGTSDALTTGNGEVSF	60		
QY 217	SDNTSSDGAIFTEASVTISNNKVSFIDNKVTGASSTTGDMGGAICAYKTSTPTKV	276		
DB 61	SDNTSSDGAIFTEASVTISNNKVSFIDNKVTGASSTTGDMGGAICAYKTSTPTKV	120		
QY 277	TLTGNOMLFSNNSTTAGGAIVYKLELASGGLTFSRNSVNGTAPKGAIAIEDSGE	336		
DB 121	TLTGNOMLFSNNSTTAGGAIVYKLELASGGLTFSRNSVNGTAPKGAIAIEDSGE	180		
QY 337	LSLSADSGDIYFLGNVTYSTTPGTRNRSIDIGTSAKMTALRSAGRAIYFPDPTTSSST	396		
DB 181	LSLSADSGDIYFLGNVTYSTTPGTRNRSIDIGTSAKMTALRSAGRAIYFPDPTTSSST	240		
QY 397	TVTQVLKVNTPPADSAIQYNGNIIFTEGKLTSETPAASKNLTSKLDPVTLSSGTTLSKH	456		
DB 241	TVTQVLKVNTPPADSAIQYNGNIIFTEGKLTSETPAASKNLTSKLDPVTLSSGTTLSKH	300		
QY 457	GVTLOTQAFTOQADSRLQEMDVGTTLEPADSTINNLVINISSIDGAKKAKIETRAITSKNL	516		
DB 301	GVTLOTQAFTOQADSRLQEMDVGTTLEPADSTINNLVINISSIDGAKKAKIETRAITSKNL	360		
QY 517	TLSTGTTLLDPTGTFYFNHSLRNQSDILELKASGVTSRAVTPDDIMGKFKHYQGT	576		
DB 361	TLSTGTTLLDPTGTFYFNHSLRNQSDILELKASGVTSRAVTPDDIMGKFKHYQGT	420		

**BEST AVAILABLE COPY**

---

```
Db 647 K--AIVGAG-----DDTVVSSATLIGAGSVNGGDTDLVANVNGSSFSADPAFGF 698
QY 406 ET--PADSALOYGNII-FTGEKLETEEA-----432
Db 699 ETLAVAGAAAGSHNANGFTMLQIGATAGATTFTNVAVNGLVLAAPGTGTTVTLANAT 758
QY 433 ---DSKMLTSKLLQPVTLISGGLTSL-----KHGVTLOTOA-----FT 466
Db 759 GTSDFVFNLT--LSSSALAAGTVALAGVETVNIATDTNTTAHVDTLTLOQTSKSIYVT 816
QY 467 QOADSRLMDVGTTLLEPADTSTI-----NNLVINISSIDAKKAKIETKATSK 514
Db 817 GNAGLNTLNTGNTAVTSEFDSASAVTGTAPAVTEVSANTTVEVVTIRGAGADSLTGSATA 876
QY 515 NLTL---SGTITLDPGTFTFENHSLRNPOSYDILEKASGTVSTAVTPDPINGEKFXH 571
Db 877 NDTIIGAGADTLVYTGCT---DFTGCGADIFDINALGTSTAFVTTIDAAVGDIDL 932
QY 572 GYOGTWGPPIWGT-GASTT 589
Db 933 VGISTNGAIDGAFGAAVT 951
```

```
RESULT 15
US-09-142-648B-7
; Sequence 7, Application US/09142648B
; Patent No. 6210948
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingie, Made H.
; APPLICANT: No. 6210948ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULI
; FILE REFERENCE: 08106/002002
; CURRENT APPLICATION NUMBER: US/09/142,648B
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: PCT/CA97/00167
; PRIOR FILING DATE: 1997-03-10
; PRIOR APPLICATION NUMBER: US 07/614,377
; PRIOR FILING DATE: 1996-03-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1026
; TYPE: PRF
; ORGANISM: Caulobacter crescentus
US-09-142-648B-7
```

Query Match 5.0%; Score 240.5; DB 4; Length 1026;  
Best Local Similarity 24.8%; Pred. No. 6.1e-10;  
Matches 183; Conservative 86; Mismatches 253; Indels 217; Gaps 35;

```
QY 12 TFAIFPLSMATETVLDSSASFDGNKNGNFSVRESQDAGTTLFKGNVTLENIP---G 67
Db 269 TFVAGEVAGAAATLVGDTLGGAGT-----DVLNMGAAAVTALPTG-VTISGIEFMNV 322
QY 68 TGTALITKSCFNNTKGDILFTGNGNSLLFQTVDA-----TVAGAAVNSSVVDKSTTFI 120
Db 323 SGAITLNTSSGVYTG-LTALNTNTSGAAGTGTAGAGCNLTATTAQAANNVAVD----- 375
QY 121 GFSLSLFIAPSGSITTKGKAVSCSTGSLKFDKNVSLFKNFTSDNGCAITAKTSLTG 180
Db 376 GRANVTYASTGVSTGTTVYGANSAASGTV---SVSV---ANSTTTTGAI-----AVTG 423
QY 181 TTMASLFESENTSSKKGAIGTSDALTFTTNGG--EVSFSDNTSSDGAALF--TEASVTI 236
Db 424 GT--AVTVAGTGAAGNAVTTTLQADVYTGSSSTTAATVYVOTAAATAGATVAGRVNGAVTI 481
QY 237 SNNV-----KVSFLDNKVTA-----SSSTTGDMG-----GAICAVKSTSTDT 274
Db 482 TDSAAASATYAGKIAATVTLGSGAATIDSALMTVNLGCTGTSLGIGALTA----- 534
```

```
QY 275 KVTLLGNOMLLFSSNNTSTAGAIYKKLELASGGLTFSPRSVNGGTAPKG-GAIAIED 333
Db 535 --TPTANTLTILNVNGLTIT--GALTDESAADDG-----FTININAGSTASTIASLVAAD 586
QY 334 SGELSLSDSGDIV-----FLGNFVST-----TPGTNRSSIDLGTS 371
Db 587 AFTLINIGDARVTTISHTAALTGITVTVNSVGATLGAELATGLVFTGAGHDSILLGATY 646
QY 372 KMTALRSAGAAIFYDPITGSSSTV-----TDVLKYN-----405
Db 647 K--AIVGAG-----DDTVVSSATLIGAGSVNGGDTDLVANVNGSSFSADPAFGF 698
QY 406 ET--PADSALOYGNII-FTGEKLETEEA-----432
Db 699 ETLAVAGAAAGSHNANGFTMLQIGATAGATTFTNVAVNGLVLAAPGTGTTVTLANAT 758
QY 433 ---DSKMLTSKLLQPVTLISGGLTSL-----KHGVTLOTOA-----FT 466
Db 759 GTSDFVFNLT--LSSSALAAGTVALAGVETVNIATDTNTTAHVDTLTLOQTSKSIYVT 816
QY 467 QOADSRLMDVGTTLLEPADTSTI-----NNLVINISSIDAKKAKIETKATSK 514
Db 817 GNAGLNTLNTGNTAVTSEFDSASAVTGTAPAVTEVSANTTVEVVTIRGAGADSLTGSATA 876
QY 515 NLTL---SGTITLDPGTFTFENHSLRNPOSYDILEKASGTVSTAVTPDPINGEKFXH 571
Db 877 NDTIIGAGADTLVYTGCT---DFTGCGADIFDINALGTSTAFVTTIDAAVGDIDL 932
QY 572 GYOGTWGPPIWGT-GASTT 589
Db 933 VGISTNGAIDGAFGAAVT 951
```

Search completed: October 2, 2001, 03:27:58  
Job time: 7240 sec

TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 415-0813  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1536 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE:  
 US-08-719-641-2

Query Match 5.1%; Score 243.5; DB 4; Length 1536;  
 Best Local Similarity 22.6%; Pred. No. 6.7e-10;  
 Matches 138; Conservative 105; Mismatches 226; Indels 141; Gaps 29;

QY 20 MATEVLDSSAFDGNKNGNSVRESQEDAGTYLFFKNTLENIPGCTLAITSCFNN 79  
 DB 818 MGKGIYARKNITFEFG-NITFGSKRAVE-----IEGNVTINN--NANVTILGSDFDN 868  
 QY 80 TKGDLT-----FTGNGNSLLFQTVADGTVAGAAYN---SSVYDKSTPFIQFSSLSFIASP 131  
 DB 869 HOKPLTIKKDVIINSGNL-----TAGNIVINAGNLIYESNANFALITNFTNVG- 918  
 QY 132 GSITTGKAVSCSTGSLKF-----DKNVSL--FSKNSTPDNGAIAITAKT--LSLT--G 180  
 DB 919 GLEFDNKGNSNISIAGKAFKIDNSKNLSITNSSSTRTIISGNTTKNDLNTNNG 978  
 QY 181 TMSALFSENTSKGGAIQTSDALIT-----GNQGVSPSDNTSSDGAIFTEAS 233  
 DB 979 SPTEMOIGDVGQKGEENLTISSDKINITKQITIKAGVDENSDSNANNTNANITIKTEIK 1038  
 QY 234 VT-----IS--NNAKVSFIDNKVTGASSSTGDMSGAICAYKTSIDTKYTLGNOMLFS 287  
 DB 1039 LTQDLNISGFNAEITAKGSDLTIGTNTSADGTNAKKVTINQVADKISADGHKVTLS 1098  
 QY 288 -----NNTSTTAGAIIYVKKLELASGTLTFSRN--SVNGGTAPKGAIAIEDSGELS 338  
 DB 1099 KYETSQGNNTEDSDN-----NAGLTIDAKNVYNNN-----ITSHKAVS 1139  
 QY 339 LSADSGDIYF-LGNVTYSTPG---TNRSSIDLG---TSAKMTALRSAGAIIYDPI 390  
 DB 1140 ISATSEITTKGTITNATGVEITTAQGSILIGIESSSGSVTLTATGALAV---SN 1195  
 QY 391 TGGSTTVDVIVKNETPPASALQYTGNIITGEKLSFEADSKNLTSKLQPTLSGG 450  
 DB 1196 ISGNTVYVANGALITLAGSTIKGESV-----TTSQSGDIG-----TISGG 1240  
 QY 451 TSLKHGVTLOAFQTQADSRLMDVGTLEPADTSTIN-----NLVIN----- 495  
 DB 1241 TVEVK-----ATESLITQSNKIKATTCGEANTYSANGFTIGTISGNTVAVTANAGDLFVG 1295  
 QY 496 -----ISSIDGA-----KKAKIETKATSKNLJLSGTTILLDPGTGFENSLANPOSYDL 546  
 DB 1296 NGAELNATGAATLTSSGKLTTEASTSHITSAGQVNLASADGSAVAGSIANA-----V 1349  
 QY 547 ELKASGTVS 556  
 DB 1350 TLNTGTTLT 1359

RESULT 14  
 US-08-614-377A-7  
 Sequence 7, Application US/08614377A  
 Patent No. 5976864  
 GENERAL INFORMATION:  
 APPLICANT: Smith, John  
 APPLICANT: Bingle, Wade H.  
 APPLICANT: NO. 5976864ellini, John F.  
 TITLE OF INVENTION: EXPRESSION AND SECRETION OF  
 TITLE OF INVENTION: HEREROLOGOUS  
 TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER  
 NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson PC  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/614,377A  
 FILING DATE: 12-MAR-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/194,290  
 FILING DATE: 09-FEB-1994  
 CLASSIFICATION: 435  
 APPLICATION NUMBER: US 07/895,367  
 FILING DATE: 09-JUNE-1992  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Tsao, Y. Rocky  
 REGISTRATION NUMBER: 34053  
 REFERENCE/DOCKET NUMBER: 08106/002001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-542-5070  
 TELEFAX: 617-542-8906  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1026 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-614-377A-7

Query Match 5.0%; Score 240.5; DB 2; Length 1026;  
 Best Local Similarity 24.8%; Pred. No. 6.1e-10;  
 Matches 183; Conservative 86; Mismatches 253; Indels 217; Gaps 35;

QY 12 FPAIPLSMIATEFVLDSSAFDGNKNGNSVRESQEDAGTYLFFKGVTLNIP-----G 67  
 DB 269 TVAAEYVAGAAITLYGDTLSGAGT-----DVLNMYQAAAVTALPTG-VTISGIEETMNT 322  
 QY 68 TGTATTKSCFNNTKCDLFTTGNGNSLLFQTVDA-----TVAGAAVNSVYDKSTPEI 120  
 DB 323 SCALITLMTSSGVGT-LTALNTNTSGAQTVAAGQNLTAATTAQAANNVAVD----- 375  
 QY 121 GSSSLFSLAPSSITITGKAVSCSTGSLKPKKNVSLFSKRFSDNGCAITAKLSTLG 180  
 DB 376 GRANTVASTGVTSGTTTGVANGASAASTV---SVSV---ANSTTTTGAT-----AVTG 423  
 QY 181 TMSALFSENTSKKGAIQTSDALITIGNOG--EVSEFSDNTSSDGAIF--TEASTYI 236  
 DB 424 GT--AVTVAQTAGNAVNTTLLADVTIGNNSSTTAVTYQTAAATAGATVAGRVGATYI 481  
 QY 237 SNNA-----KVSFIDNKVTGA-----SSSTGDMSG-----GAICAYTSTDT 274  
 DB 482 TDSAAASATTAACKIATVTLLGSEGAFTIDSALTIVNLSTGTSLSIGRGALTA----- 534  
 QY 275 KYTLGNOMLFSNNTSTTGAIIYVKKLELASGTLTFSRNSVNGTAPKG-GAIAIED 333  
 DB 535 --TTPANTLLTNVNGLTFT--GAITDSEAADG---FTTININGSSTAASSTIASLVAD 586  
 QY 334 SGEELSDSGDIY-----FLGNTVTS-----TPGNRSSIDLGTS 371  
 DB 587 ATTLLISGDARTYTSHTAALTLGTTVINSVATLGAELATGLVETGGAGRDSIILGAT 646  
 QY 372 KMTALRSAGAIIYDPIITGSSSTTV-----TDVLKVN----- 405



Db 1296 NGAINTEGAALITSSGKLTTEASSHTSAKGOVNLASQDGSVAGSINANAN-----V 1349  
QY 547 ELKASGTVTS 556  
Db 1350 TLNTGTGLT 1359

RESULT 12  
US-08-617-697-2  
Sequence 2, Application US/08617697  
Patent No. 5977336  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/617,697  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 05-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-557  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-617-697-2

Query Match 5.1%; Score 243.5; DB 2; Length 1536;  
Best Local Similarity 22.6%; Pred. No. 6.7e-10;  
Matches 138; Conservative 105; Mismatches 226; Indels 141; Gaps 29;  
QY 20 MATEVLDSSAPDGKKNKNGFSVRESQEDAGTYLFKGNVTLEINGGTGATIKSCFN 79  
Db 818 MICKGIYAKKNITFEFG-NITFGSKAVTE-----TEGNVTINN--NANTVLIGSDFDN 868  
QY 80 TKGDLT-----FTGNGNSLTFQYDAGTVAGAVN---SSVVDKSTTFIGSSLSFTASP 131  
Db 869 HQPKPLTIKKVVIINSGL-----TAGNIVINAGNLTVESNAFKAITNFTFVNG- 918  
QY 132 GSSITTKGAVSCSTGSLK-----DKNVSL--FSKNFSTNGCAITPAAT--LSLT--G 180  
Db 919 GLPDNKGNSNISIAKGARPKIDNSKNLSITNNSSTYRTIISGNTNKNKNDLNTNEG 978

QY 181 TTMALEFSENTSSKKGALOTSDALIT-----GNOGEVSPSDNNTSPDGAALFEAS 233  
Db 979 SDEMOIGGDVSOKEGNULTISSDKINITKOTIKAGVDGENSDSDATNNANLITKTELK 1038  
QY 234 VT-----IS--NNAVFSTIDNKVKTGASSSTTGDMGGAICAKTISTDKRVLTGNQMLFS 287  
Db 1039 LTODLNISGFNKAETFAKDSGLTIGNTNSADCTNAKRVFNQYKDSKISADGKRVLHS 1098  
QY 288 -----NNTSTAGAGAIYKKLELASGGLTFSRN--SVNGGTAPRPGALAEIDSELS 338  
Db 1099 KVETSGSNNTEDSSDN-----NAGLTIDAKVTYNNN-----ITSKRAVS 1139  
QY 339 LSADSGDIVF-IGNTVNTTPG---TNRSSIDLG---TSAKMTALRSAGRAIFYDPI 390  
Db 1140 ISATSGEITTTGTITNATGCVNVEITAGTSLIGISSSGSVLTLTEGALAN-----SN 1195  
QY 391 TTGSSVTVDLVKNEFPADSALOYTGNIIFTGKISLSETEADSKNLTKLLOPVTLSGC 450  
Db 1196 ISGNTVVTANSALTLTLAGSTIKGTESV-----TSSQSGDIGG-----TISGC 1240  
QY 451 TLSLKHGVTLOTQAFTOQADSRLEMDVGTILPADTSTIN-----NLVIN----- 495  
Db 1241 TVEVK-----ATESLTQNSKIKATTEANVTATGTIGTISGNTVNTANAGDYLTVG 1295  
QY 496 -----ISSIDGA-----KKAKETKATSKNLTLSGTLTLPDPTGFYENHSLRNPOSYDIL 546  
Db 1296 NGAINTEGAALITSSGKLTTEASSHTSAKGOVNLASQDGSVAGSINANAN-----V 1349  
QY 547 ELKASGTVTS 556  
Db 1350 TLNTGTGLT 1359

RESULT 13  
US-08-719-641-2  
Sequence 2, Application US/08719641  
Patent No. 6218141  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/719,641  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-625  
TELECOMMUNICATION INFORMATION:

MOLECULE TYPE: DNA (genomic)  
US-08-469-880-2

Query Match 5.1%; Score 243.5; DB 2; Length 1536;  
Best Local Similarity 22.6%; Pred. No. 6.7e-10;  
Matches 138; Conservative 105; Mismatches 226; Indels 141; Gaps 29;

20 MATEVYLDSSAFDKNKGNFVSRESOEDACTTFLKGNVTLNIPGTGATKSCFNN 79  
DB 818 MICKGIYAKKNITFEKG-NITFGSKRAVTE-----IEGNVTINN--NANVTILGSDFDN 868  
QY 80 TKGDLT-----FTGNGNSLLFOTVDAGTYAGAAYN--SSVYDKSTTFEGFSSLSFIAP 131  
DB 869 HOKPLIKKDVIIINSGL-----TAGGNVTNAGNLTVESNANFPAITNFTFNVC- 918  
QY 132 GSSITTKGKAVSGSTGSLK-----DKNVSLL--FSKNFTDNGALITAKT--LSLT--G 180  
DB 919 GLEPNKGNNSIYAKGAFKIDNSKNLSITNNSSTYRTIISGNITPKNGDLNITNEG 978  
QY 181 TTMSALFSENTSSKKGAIOTSDALIT-----GNOGEVFSMDTSSDSCGAITFEAS 233  
DB 979 SDTEMOIGGDVQSKENGLTISDKINITKOITTKAGVDENSDSDATNNANLITKTELK 1038  
QY 234 VT-----IS--NNAKVSFIDNKVTGASSSTTGDMGSGAICAYKTSTDYKVTLGNOMLFS 287  
DB 1039 LFODLNIGFNKAEITAKGSDLTIGNTNSADGTNAKKVTFNQVKDSKISADGKVTLS 1098  
QY 288 -----NNTSTTAGGAIYVKLELASGGLTLEFSN--SYNGSTAKGGAIAIEDSGELS 338  
DB 1099 KVEITSGSNNTEDSDN-----NAGLTIDAKNVYNNN-----ITSKRAVS 1139  
QY 339 LSADSGDIVE-LGNVTSTTPG---TNRSSIDG---TSAKMTALRSAAGRAIYFYDPI 390  
DB 1140 ISATSEITTKGTITNATTGNVEITAOGSLIGIESSSGSVTLTATGALAV-----SN 1195  
QY 391 TTGSSSTTVTVKNETPADSALOYTGNIITFGEKLETFEADSKNLSKLLQPYTLSCG 450  
DB 1196 ISGNVTYVANSAGALTTLTGSTIKGTESV-----TTSSQSGDIG- -TISGG 1240  
QY 451 TSLKRGVLTQOAFQOADSRLKEMOVGTTLEPADTSTIN-----NLVIN----- 495  
DB 1241 TVEVK-----ATESLTTQNSKIKATGGEANVTSAITGIGTISGNTVAVTANAGDLTVC 1295  
QY 496 -----ISSIDGA-----KKAKIETKATSKNLTLSGTTLLDPTGTFEENSLRNPOSYDIL 546  
DB 1296 NGAEINATEGAALITSSGKITTEASHITSAKQVNLNADGSAVGSINMAN-----V 1349  
QY 547 ELKASGTVTS 556  
DB 1350 TLNVTGTLTT 1359

RESULT 11  
US-08-728-470-2  
Sequence 2, Application us/08728470  
Patent No. 5928651  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OR INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: Of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,470  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:  
US-08-728-470-2

Query Match 5.1%; Score 243.5; DB 2; Length 1536;  
Best Local Similarity 22.6%; Pred. No. 6.7e-10;  
Matches 138; Conservative 105; Mismatches 226; Indels 141; Gaps 29;

20 MATEVYLDSSAFDKNKGNFVSRESOEDACTTFLKGNVTLNIPGTGATKSCFNN 79  
DB 818 MICKGIYAKKNITFEKG-NITFGSKRAVTE-----IEGNVTINN--NANVTILGSDFDN 868  
QY 80 TKGDLT-----FTGNGNSLLFOTVDAGTYAGAAYN--SSVYDKSTTFEGFSSLSFIAP 131  
DB 869 HOKPLIKKDVIIINSGL-----TAGGNVTNAGNLTVESNANFPAITNFTFNVC- 918  
QY 132 GSSITTKGKAVSGSTGSLK-----DKNVSLL--FSKNFTDNGALITAKT--LSLT--G 180  
DB 919 GLEPNKGNNSIYAKGAFKIDNSKNLSITNNSSTYRTIISGNITPKNGDLNITNEG 978  
QY 181 TTMSALFSENTSSKKGAIOTSDALIT-----GNOGEVFSMDTSSDSCGAITFEAS 233  
DB 979 SDTEMOIGGDVQSKENGLTISDKINITKOITTKAGVDENSDSDATNNANLITKTELK 1038  
QY 234 VT-----IS--NNAKVSFIDNKVTGASSSTTGDMGSGAICAYKTSTDYKVTLGNOMLFS 287  
DB 1039 LFODLNIGFNKAEITAKGSDLTIGNTNSADGTNAKKVTFNQVKDSKISADGKVTLS 1098  
QY 288 -----NNTSTTAGGAIYVKLELASGGLTLEFSN--SYNGSTAKGGAIAIEDSGELS 338  
DB 1099 KVEITSGSNNTEDSDN-----NAGLTIDAKNVYNNN-----ITSKRAVS 1139  
QY 339 LSADSGDIVE-LGNVTSTTPG---TNRSSIDG---TSAKMTALRSAAGRAIYFYDPI 390  
DB 1140 ISATSEITTKGTITNATTGNVEITAOGSLIGIESSSGSVTLTATGALAV-----SN 1195  
QY 391 TTGSSSTTVTVKNETPADSALOYTGNIITFGEKLETFEADSKNLSKLLQPYTLSCG 450  
DB 1196 ISGNVTYVANSAGALTTLTGSTIKGTESV-----TTSSQSGDIG- -TISGG 1240  
QY 451 TSLKRGVLTQOAFQOADSRLKEMOVGTTLEPADTSTIN-----NLVIN----- 495  
DB 1241 TVEVK-----ATESLTTQNSKIKATGGEANVTSAITGIGTISGNTVAVTANAGDLTVC 1295  
QY 496 -----ISSIDGA-----KKAKIETKATSKNLTLSGTTLLDPTGTFEENSLRNPOSYDIL 546

QY	547	ELKASGTVTS	556
	1	:  : :	
Db	1350	TLNTTGTLLT	1359

RESULT 9

US-08-530-198-2  
; Sequence 2, Application US/08530198

; GENERAL INFORMATION:

APPLICANT: BARENKAMP, STEPHEN J  
APPLICANT: ST. GEME III, JOSEPH W

TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
OF NON-TYPEABLE HAEMOPHILUS

```

; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
;

```

ADDRESSEE: Shoemaker and Mattare, Ltd  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1  
CITY: Arlington

```

QY 288 -----NNTSTAGAIYVKKLELASGGLTFEN-N-SYNGCTAPKGAIAIEDSGELS 338
Db 1099 KVETSGNNNTTEDSSDN-----NAGLTLDANAVVNN-----TISHKAVS 1139
QY 339 LSAADSGIIV-LGNVYVSTPG--TNRSSIDG-----TSAKMFLRSAGRAIFYDPI 390
Db 1140 ISATSGIITTKTGTTINATTGANNETIAOQGISLIGIESSSGSVTLTANEGALAV-----SN 1195
QY 391 TTGSGSVTVDLKVNETPADSALOYTGNIIFGKLESTPADSKNLTSKLOPVLISGG 450
Db 1196 ISGNVTVFVANSGLVTLTLAGSTIKGIEV-----TTSQSGDIDG-----TISGG 1240
QY 451 TYSLKHSVTLQOTAFQOADSRLMDVGYTLLEPADTSTIN-----NLVIN----- 495
Db 1241 TVEYK-----ATESLTPQSSKIKATGTGANVTSATGTIGCTISGNFVNVYVANNAGDLTVG 1295
QY 496 -----ISSIDA-----KKATIEKATSKNULTLSGTLTLLDPGTGYEYHSLRNPOSYIL 546
Db 1296 NGAEINNTTEGAANLTTSSGSLTTEASHSTISAKGOVNLSADQSVAGSINAN-----V 1349
QY 547 ELKASGVTS 556
Db 1350 TLNNTGTPLT 1359

```

RESULT 10  
US-08-469-880-2

```

; Sequence 2, Application
; Patent No. 5876733
; GENERAL INFORMATION

```

APPLICANT: Barenkamp, Stephen J. Weight Surface Proteins  
 TITLE OF INVENTION: High Molecular Weight Haemophilus  
 TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Shoemaker and Mattare, Ltd  
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: 5109. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469, 880  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302, 832  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-038-682-2

Query Match 5.1%; Score 243.5; DB 1; Length 1536;  
Best Local Similarity 22.6%; Pred. No. 6.7e-10;  
Matches 138; Conservative 105; Mismatches 226; Indels 141; Gaps 29;

```

QY 20 MATETVLDSSAFPGNKNNGNFSVRESQEDACTYTLFKGNVLENIPTGTAITSCFPN 79
DB 818 MGKGIIVAKKNITFEFG-NITFGSKRAVTE-----ISGNVTINN--NANVTLLISDSDN 868
QY 80 TKGDLT-----FTGNGNSLIFQTVAGYVAGAANV---SSVVDKSTTFEGFSLFIASP 131
DB 869 HOKPLTIKKDVIYNSGNL-----TAGNVIYIAGNLIVESANFKATINFPFNNV- 918
QY 133 GSSITTTGKAVSCSTGSLKF-----DKNVSLL--FSKNFSTNDGATIAKT--LSLT--G 180
DB 919 GLFDKNGNSNISIAKGARFKDIDNSKNLSITTNSSSTYRTIISGNITKNKNDLNTMEG 978
QY 181 TTMASLFSFENSSKKGAIGTSDALIT-----GNGEVSFSNTSSDSGAALFTFAS 233
DB 979 SDPEMOIGADVQKQKGNLTISDDKINITKQITIKAGVDGENDSDATNNANLTIKTELK 1038
QY 234 VT-----IS--NNAKVSFDNKNVTGASSSTTGDMSCAICAYTSTDFKVTLLGNOMLFS 287
DB 1039 LTQDLNISGFNAELTADGSDLTIGNTNSADGTNAKYTFNOVADSKISADGHKVTLLHS 1098
QY 288 -----NNTSTAGGAIYVKLELBSGGLTLFSRN--SVNGTAPKGAIALIEDSGELS 338
DB 1099 KVEITSGSNNTFEDSSDN-----NAGLTIDAKNVTYNNN-----ITSHKAVS 1139
QY 339 LSADSGDIYF--LGNVTYSTTPG---TNRSSIDLG---TSKMTALRSAGAIIFYDPI 390
DB 1140 ISATSGEITTKGTTINATGVEITTAQGTSLGIESSSGVVTLTAEGALAV-----SN 1195
QY 391 TTGSSSTTVTVLVKNEFPADSAIQTGNIIFTEGKILSEFADSKNLTKLQPTLSGG 450
DB 1196 ISGNVTYVYANGALTTLAGSTIKTESV-----TTSQSDDIG-----TISGG 1240
QY 451 TLSLKHGVTLOAFTQAADSRLEMDVGTLEPADTSTIN-----NLVIN----- 495
DB 1241 TVEVK-----ATESLFTOSNSKIKATTEGAVNTSATGTIGTISGNTVAVTANAGDLTVG 1295
QY 496 -----ISSIDGA-----KKAKIETKATSKNLTLSGTITLLDPTGTFEYHNSLRNPSTYL 546
DB 1296 NGAETNATEGAATLTTSQGLTTEASSHITTSKAGOVNLSAODGSVAGSINAA-----V 1349
QY 547 ELKASGTVVS 556
DB 1350 TLNTGTTLTT 1359

```

RESULT 8  
US-08-302-832-2  
Sequence 2, Application US/08302832  
Patent No. 5603938

GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Wattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/302,832  
FILING DATE: 16-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US pct/us93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berksstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-404  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-302-832-2

Query Match 5.1%; Score 243.5; DB 1; Length 1536;  
Best Local Similarity 22.6%; Pred. No. 6.7e-10;  
Matches 138; Conservative 105; Mismatches 226; Indels 141; Gaps 29;

```

QY 20 MATETVLDSSAFPGNKNNGNFSVRESQEDACTYTLFKGNVLENIPTGTAITSCFPN 79
DB 818 MGKGIIVAKKNITFEFG-NITFGSKRAVTE-----ISGNVTINN--NANVTLLISDSDN 868
QY 80 TKGDLT-----FTGNGNSLIFQTVAGYVAGAANV---SSVVDKSTTFEGFSLFIASP 131
DB 869 HOKPLTIKKDVIYNSGNL-----TAGNVIYIAGNLIVESANFKATINFPFNNV- 918
QY 133 GSSITTTGKAVSCSTGSLKF-----DKNVSLL--FSKNFSTNDGATIAKT--LSLT--G 180
DB 919 GLFDKNGNSNISIAKGARFKDIDNSKNLSITTNSSSTYRTIISGNITKNKNDLNTMEG 978
QY 181 TTMASLFSFENSSKKGAIGTSDALIT-----GNGEVSFSNTSSDSGAALFTFAS 233
DB 979 SDPEMOIGADVQKQKGNLTISDDKINITKQITIKAGVDGENDSDATNNANLTIKTELK 1038
QY 234 VT-----IS--NNAKVSFDNKNVTGASSSTTGDMSCAICAYTSTDFKVTLLGNOMLFS 287
DB 1039 LTQDLNISGFNAELTADGSDLTIGNTNSADGTNAKYTFNOVADSKISADGHKVTLLHS 1098
QY 288 -----NNTSTAGGAIYVKLELBSGGLTLFSRN--SVNGTAPKGAIALIEDSGELS 338
DB 1099 KVEITSGSNNTFEDSSDN-----NAGLTIDAKNVTYNNN-----ITSHKAVS 1139
QY 339 LSADSGDIYF--LGNVTYSTTPG---TNRSSIDLG---TSKMTALRSAGAIIFYDPI 390
DB 1140 ISATSGEITTKGTTINATGVEITTAQGTSLGIESSSGVVTLTAEGALAV-----SN 1195
QY 391 TTGSSSTTVTVLVKNEFPADSAIQTGNIIFTEGKILSEFADSKNLTKLQPTLSGG 450
DB 1196 ISGNVTYVYANGALTTLAGSTIKTESV-----TTSQSDDIG-----TISGG 1240
QY 451 TLSLKHGVTLOAFTQAADSRLEMDVGTLEPADTSTIN-----NLVIN----- 495
DB 1241 TVEVK-----ATESLFTOSNSKIKATTEGAVNTSATGTIGTISGNTVAVTANAGDLTVG 1295
QY 496 -----ISSIDGA-----KKAKIETKATSKNLTLSGTITLLDPTGTFEYHNSLRNPSTYL 546
DB 1296 NGAETNATEGAATLTTSQGLTTEASSHITTSKAGOVNLSAODGSVAGSINAA-----V 1349

```

```

Oy 451 --TSLKH-----GVTIDTO-AFGQADSRLEMDVG-TLEPAD-----TSLIN 491
Db 1322 QTTILFADSSLAGINANAANTLTGTLITGTSKRNATASGLITLIINADAKIDGASGR 1381
Oy 492 LYVNISSIDGAKKAKITKATSKNULSLGTYITLLDPTGYFFEN 534
Db 1382 TVVNATNAGSGNV--TATSSSVNTGLNLTINSLNTISEN 1421

```

```

RESULT      6
US-08-617-697-10
Sequence 10, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
of Hemophilus
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shemmaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-10

```

Query Match	5.7%;	Score 272;	DB 2;	Length 1600;
Best Local Similarity	22.7%;	Pred. No. 4.7e-12;		
Matches 147;	Conservative 76;	Mismatches 218;	Indels 206;	Gaps 29;

[illegible][illegible]

```

RESULT 7
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
APPLICANT: BAREKAMP, STEPHEN J
APPLICANT: ST. GENE ILL, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINSS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shemmaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BEKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-728-470-10

```

Query Match	5.7%	Score 273;	DB 2;	Length 1529;
Best Local Similarity	22.7%;	Pred. No. 3.7e-12;		
Matches 146;	Conservative 77;	Mismatches 218;	Indels 202;	Gaps 30

QY	27	LDSSAFDQKKNNEFVSRES----	QODAGTYFLFKGNVLU-ENIGPOTGAIRKSCFNMTK	81	
Dd	846	LNTGTFITNNGTANINIKQGVVLQGDINN----	KGLNITTINASTQCTIINGNITHEK	901	
QY	82	GDLTFT-----GN-----	GNSLFFQ-----	TYDACTVACAANVSVVDKS	116
Dd	902	GDILINKIKADAEIOIGGNISQKREMLITSSDKVNITNQITIKAGEGGRSDSSSEAMEN		961	
QY	117	TT-----	PIGESSLFISPGSSITTGCA-VSCSTGSIKRDK-	153	
Dd	962	LTTIOFKELKLAGLNIISGEFKAEITAKKNSDILLTGNAISGNDARKVYFDFKDKDSTID		1021	
QY	154	--NVSLEFSKNESTDNGCAITAKTTLISLPTGMSA--	LESEMTSSKK-----	GAIOYT	201
Dd	1022	GHNVTL--NEVYTSNGSS-NGNONSSTGLTISAKVYVNNNVTSHKRTINISAAGNYTT		1078	
QY	202	SDALTTIGNOGEVFSFSDNVSDDSGAIFTEASYTTSINNAKVSFIDNKYTGASSST----		256	
Dd	1079	KEGTTINATTGSEVVAQONCTIGN--	ITSQNTVATENLVTTENAVINATSGVNVIST	1136	
QY	257	-TGMGSGCAICAKTSTDNFTVLTKNOMLFSNNT-----	STTGGAIAIYK-	301	
Dd	1137	KTGDIKG---ISTSGNNVITASGN-TLKVSNIITGQDVTYVADAGALTTAGSTISATY		1192	
QY	302	-----	KIELASGGLTFSRNSVNGTAIPKGAIAI-EDSGEL-----	337	
Dd	1193	GNANITTKTDINGKVESSGSVTLVATGATLAVGNISGNTVITITADSGKLTFSVGTIN		1252	
QY	338	-----SLSADSGNI--	VPJGNVTYTTPTGNSSIDIGSAAKMTLRSAAGAIIFYDPI	390	
Dd	1253	GTNSVYTSOSGDIECTISGNTVNYVT--	ASTGDILLTASNAVEAKNGA-----	1295	
QY	391	TTGSSYTVTVLVLENTPADSALOYGNITTEGKLESEAEADSKNLTSLQPYTSLSG		450	
Dd	1300	-----	TLTAEISKLTQTIGSSITSSNG	1322	
QY	451	--TISLKH-----	GVTLOTO-AFPQOADSRLMDVGT-TLEPAD-----	TSTINN	491
Dd	1322	QTYTLTADSSIGAINANAVYTLTTTGTTLTTTGDSKINATSGTLTITNAKPAKLDGAASGDR		1381	
QY	492	LVINISSIDGAKAKIETKATSKNLTSLGTITLIDTGTGFYFN		534	
Dd	1382	TYVNAVITASGCVN--	TAKTSSVAVITGDNLNTINGNLINISEN	1421	

RESULT 5  
 US-08-719-641-10  
 : Sequence 10, Application US/08719641  
 : Patent No. 6218141  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Barenkamp, Stephen J  
 : TITLE OF INVENTION: High Molecular Weight Surface Proteins  
 : NUMBER OF INVENTION: of No. 6218141-Typeable Haemophilus  
 : NUMBER OF SEQUENCES: 10  
 :  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Shoemaker and Mattare, Ltd.  
 : STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 : STREET: Bldg. 1  
 : CITY: Arlington  
 : STATE: Virginia  
 : COUNTRY: U.S.A.  
 : ZIP: 22202-0286  
 :  
 : COMPUTER READABLE FORM:

MEDIA TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

Query Match	5.7%;	Score 273;	DB 4;	Length 1529;
Best Local Similarity	22.7%;	Pred. No. 3.7e-12;		
Matches 146;	Conservative 77;	Mismatches 218;	Indels 202;	Gaps 30;

[illegible]

Db 1190 RTVYVATNATNAGSGNV---TAKTSSSVNITGDLNTINGNLINISEN 1230

## RESULT 3

US-08-617-697-9  
; Sequence 9, Application US/08617697  
; Patent No. 5977336  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: Of No. 5977336-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/617,697  
; FILING DATE: 01-APR-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 05-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstesser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-557  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1599 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-617-697-9

Query Match 5.8%; Score 277.5; DB 2; Length 1599;  
Best Local Similarity 22.7%; Pred. No. 1,8e-12;  
Matches 146; Conservative 82; Mismatches 219; Indels 197; Gaps 29;  
QY 23 TETVLDSSAFDGNKNGNFVRES-----OEDAGTYY--LFKGVNTLENI 65  
Db 912 TMTFNWAGSFDDNNGANSIARAGAKFKDINNNTSSNITTNTSTTTRTIKGNIS--NK 969  
QY 66 PGTGTATKSC-----FNNTKGDLFTFGNGNSLLFO--TYDAGTVAGAAVSSVYDK 115  
Db 970 SGLDNIIDKKSDAEIQIGNISOKEGNLTISDKVNITNTQITIKAGVEGGSDSEAEANA 1029  
QY 116 STT-----FIGSSLSFIASPGSSITTKGA--VSCSTGSLAKDK-----153  
Db 1030 NLTIQTEKLADLNTISGNKAEITAKNGSDLTIGNASGNDADKVFTEPKVDSKIST 1089  
QY 154 ---NSLLFKNSTONGATKATKTLISGTTMSA---LFSENTSSKK-----GGAIQ 200  
Db 1090 DGHNVTL--NSEYKTSIGSS--NAGNDNSTGLTISAKDYVYNNNTSKTINISAAAGNVT 1146  
QY 201 TSDALITGNGEVSFSDNTSSDSGAIFTEASVTISNNAKVSFIDNKVTGASSST----256

Db 1147 TREGTINATGSEVVAQNGTIKGN--ITSQNTVATENLVTTENAVINATSGVNIIS 1204  
QY 257 --TGDMSGATCAKKTSTDTFVTLGNOMLIFSNT-----STTAGCAIYVK 301  
Db 1205 TKTDIDIRGG--IESTGNVNTIASGN--TEKVSNTIGODVTYADAGALTATAGSTISAT 1260  
QY 302 -----KLEASGGLTFESRNSVNGGTAPKGAIAI--EDSGEL-----337  
Db 1261 TGNANITTKTDINGKYESSGSVTIVATGATLAVGNISGTVITITDSDGLSTVSGSTI 1320  
QY 338 -----SLSADSGDI--VELGNVTSTPGTNRSSIDLGTSAKKTALBSAAGRAIYFYP 389  
Db 1321 NGTNSVTTSSQSDIEGTISGNFVNVV---ASTGDLTIGNSAKVEARKGA-----1368  
QY 390 ITGSSSTTVYDVAKVNEFPADSAQTYGNITFTGEKISFEPAADSKNLTSLQPYTLSS 449  
Db 1369 -----TLTASGKLTYYTGSSITSSN 1389  
QY 450 G--TLSEKH-----GVTLOTQ--AFTQADSRLMDVGT--TLEPAD-----TSTIN 490  
Db 1390 GQTLTLAKDSSIAGNINANANTLTGTLTGTGDSKINATSGTLTINAKADLGAASGD 1449  
QY 491 NLVINISSIDGAKKAKIETKATSKNLTLSGTTLLDPTGFYEN 534  
Db 1450 RTVYVATNATNAGSGNV---TAKTSSSVNITGDLNTINGNLINISEN 1490

## RESULT 4

US-08-728-470-10  
; Sequence 10, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: Of No. 5928651-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728,470  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstesser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-633  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1529 amino acids



Query Match 5.8%; Score 277.5; DB 2; Length 1338;  
Best Local Similarity 22.7%; Pred. No. 1.3e-12;  
Matches 146; Conservative 82; Mismatches 219; Indels 197; Gaps 29;

QY 23 TETVLDSSAFDGNKNGNFVRES-----QEDAGTTY--LFKGNVLENI 65  
DB 652 TWYTFVAVAGSPDNNGASNISIRGAKFKDINNTSSLTNTTSOTYRIRIINGNIS--NK 709  
QY 66 PGTGAITKSC-----FNTRKGDLLFTGNGNSLFO-TVDACTVAGAANSSVVDK 115  
DB 710 SCDLNIIDKKSDAEIIGIGNISQKEGNLTISDKVNTNQITIKAGVEGGRDSSAEANA 769  
QY 116 SPT-----FIQSSLSFIASPGSSITTKGA--VCSSTGSLKFDK----- 153  
DB 770 NLTIQTKELKLAGDLNISGFENKAETAKNGSDLTIGNASGNAADAKVYFDVKOSKIST 829  
QY 154 ---NWSLFSKNFSTDNCAITAKTLSLGTGTMAS--LFSBNTSSK-----GGAIO 200  
DB 830 DGHNVTL--NSEVKTISNGSS--NAGNDNSTGLTISAKDVYVNNNVTSHTKINISAAAGNT 886  
QY 201 TSDALITGNOGEVSFSDNTSSDGAIFTEASVTISNNAKVSFIDNKVTGASSST---- 256  
DB 887 TREGTTINATTSVEVTAONGTIKGN--ITSQNVTVTATENLVTEENAVINATSGTVNIS 944  
QY 257 --TGMSGGAICAYKSTDTKYTLTGNOMLFSSNNM-----STTAGCAIYK 301  
DB 945 TKTGDIKGG--IESISGVNNTASGN--TLKVSNTIGQDVYVADAGALTTAGSTISAT 1000  
QY 302 -----KLEASGGLTLFSSRNVSNGTAPKGAIAI-EDSGEL----- 337  
DB 1001 TGNANITTKTGINGKVESSGSVTLVATGATLAVGNISGNTVYVADAGALTTAGSTISAT 1060  
QY 338 -----SLSADGDI--VFLGNTVSTTPGTNRSSIDLTGSAKMTALRSAGRAIYFYD 389  
DB 1061 NCTNSVTTSSOSGDIEGTISGNTVNT--ASTGDLTISNKAKEAKNGAA----- 1108  
QY 390 ITTGSSTVTVLAKVNEPADSALOYTGNIIPTGKLTSETEADSKNLTSLKLOPYTLISG 449  
DB 1109 -----TLTAESGKLTTOGSSSTTSN 1129  
QY 450 G--TSLKH-----GVTLQTO-AFTQOADSRLNDVGT-TLEPAD-----TSTIN 490  
DB 1130 GGTTLTAKOSSIAGININAMVTLNTTGLTTTGDSKINNTSGTLTINAKDALDGAASGD 1189  
QY 491 NLVINISSIDGAKKAKIETKATSKNLTLSGTTLLDPTGTFFEN 534  
DB 1190 KRVVATNATNASGSGNV---TAKTSSVNTIGDINTINGLNIISEN 1230

RESULT 2  
US-08-719-641-9  
; Sequence 9, Application US/08719641  
; Patent No. 6218141  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/719,641  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkleesser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-625  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1338 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-719-641-9

Query Match 5.8%; Score 277.5; DB 4; Length 1338;  
Best Local Similarity 22.7%; Pred. No. 1.3e-12;  
Matches 146; Conservative 82; Mismatches 219; Indels 197; Gaps 29;

QY 23 TETVLDSSAFDGNKNGNFVRES-----QEDAGTTY--LFKGNVLENI 65  
DB 652 TWYTFVAVAGSPDNNGASNISIRGAKFKDINNTSSLTNTTSOTYRIRIINGNIS--NK 709  
QY 66 PGTGAITKSC-----FNTRKGDLLFTGNGNSLFO-TVDACTVAGAANSSVVDK 115  
DB 710 SCDLNIIDKKSDAEIIGIGNISQKEGNLTISDKVNTNQITIKAGVEGGRDSSAEANA 769  
QY 116 SPT-----FIQSSLSFIASPGSSITTKGA--VCSSTGSLKFDK----- 153  
DB 770 NLTIQTKELKLAGDLNISGFENKAETAKNGSDLTIGNASGNAADAKVYFDVKOSKIST 829  
QY 154 ---NWSLFSKNFSTDNCAITAKTLSLGTGTMAS--LFSBNTSSK-----GGAIO 200  
DB 830 DGHNVTL--NSEVKTISNGSS--NAGNDNSTGLTISAKDVYVNNNVTSHTKINISAAAGNT 886  
QY 201 TSDALITGNOGEVSFSDNTSSDGAIFTEASVTISNNAKVSFIDNKVTGASSST---- 256  
DB 887 TREGTTINATTSVEVTAONGTIKGN--ITSQNVTVTATENLVTEENAVINATSGTVNIS 944  
QY 257 --TGMSGGAICAYKSTDTKYTLTGNOMLFSSNNM-----STTAGCAIYK 301  
DB 945 TKTGDIKGG--IESISGVNNTASGN--TLKVSNTIGQDVYVADAGALTTAGSTISAT 1000  
QY 302 -----KLEASGGLTLFSSRNVSNGTAPKGAIAI-EDSGEL----- 337  
DB 1001 TGNANITTKTGINGKVESSGSVTLVATGATLAVGNISGNTVYVADAGALTTAGSTISAT 1060  
QY 338 -----SLSADGDI--VFLGNTVSTTPGTNRSSIDLTGSAKMTALRSAGRAIYFYD 389  
DB 1061 NCTNSVTTSSOSGDIEGTISGNTVNT--ASTGDLTISNKAKEAKNGAA----- 1108  
QY 390 ITTGSSTVTVLAKVNEPADSALOYTGNIIPTGKLTSETEADSKNLTSLKLOPYTLISG 449  
DB 1109 -----TLTAESGKLTTOGSSSTTSN 1129  
QY 450 G--TSLKH-----GVTLQTO-AFTQOADSRLNDVGT-TLEPAD-----TSTIN 490  
DB 1130 GGTTLTAKOSSIAGININAMVTLNTTGLTTTGDSKINNTSGTLTINAKDALDGAASGD 1189  
QY 491 NLVINISSIDGAKKAKIETKATSKNLTLSGTTLLDPTGTFFEN 534

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 2, 2001, 03:27:47 ; Search time 52.09 Seconds  
(Without alignments)  
366.824 Million cell updates/sec

Title: US-09-446-677b-10  
Perfect score: 4778  
Sequence: 1 MKSFPKVFSTFAIFPLSM.....FELRGSSRNVDLAKYQF 928

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2.6/prodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2.6/prodata/2/1aa/5A.COMB.pep:\*  
3: /cgn2.6/prodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2.6/prodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2.6/prodata/2/1aa/PCPUS.COMB.pep:\*  
6: /cgn2.6/prodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	277.5	5.8	1338	2	US-08-728-470-9
2	277.5	5.8	1338	4	US-08-719-641-9
3	277.5	5.8	1599	2	US-08-617-697-9
4	273	5.7	1529	4	US-08-728-470-10
5	273	5.7	1529	4	US-08-719-641-10
6	272	5.7	1600	2	US-08-617-697-10
7	243.5	5.1	1536	1	US-08-038-682-2
8	243.5	5.1	1536	1	US-08-302-832-2
9	243.5	5.1	1536	2	US-08-530-198-2
10	243.5	5.1	1536	2	US-08-469-880-2
11	243.5	5.1	1536	2	US-08-728-470-2
12	243.5	5.1	1536	2	US-08-617-697-2
13	243.5	5.1	1536	4	US-08-719-641-2
14	240.5	5.0	1026	4	US-08-614-377A-7
15	240.5	5.0	1026	4	US-08-142-648B-7
16	234.5	4.9	1026	1	US-08-194-290-7
17	230.5	4.8	1477	1	US-08-038-682-4
18	230.5	4.8	1477	1	US-08-302-832-4
19	230.5	4.8	1477	2	US-08-530-198-4
20	230.5	4.8	1477	2	US-08-469-880-4
21	230.5	4.8	1477	2	US-08-728-470-4
22	230.5	4.8	1477	2	US-08-617-697-4
23	230.5	4.8	1477	4	US-08-719-641-4
24	225	4.7	1612	1	US-08-169-927-2
25	207	4.3	1912	1	US-08-409-995-4
26	207	4.3	1912	3	US-08-685-467-4
27	207	4.3	2353	4	US-09-377-155-33

28	207	4.3	2353	4	US-08-913-942-4	Sequence 4, Appl1
29	199.5	4.2	2123	4	US-08-968-685A-10	Sequence 10, Appl1
30	195.5	4.1	674	1	US-08-317-822A-3	Sequence 3, Appl1
31	195.5	4.1	674	1	US-08-439-818A-3	Sequence 3, Appl1
32	195.5	4.1	674	2	US-08-751-965-3	Sequence 3, Appl1
33	195.5	4.1	674	2	US-08-738-975-3	Sequence 3, Appl1
34	195.5	4.1	674	3	US-08-728-626-3	Sequence 3, Appl1
35	195.5	4.1	674	3	US-08-808-599A-3	Sequence 3, Appl1
36	195.5	4.1	1160	3	US-08-808-599A-24	Sequence 24, Appl1
37	193.5	4.0	749	1	US-08-317-822A-2	Sequence 2, Appl1
38	193.5	4.0	749	1	US-08-439-818A-2	Sequence 2, Appl1
39	193.5	4.0	749	2	US-08-751-965-2	Sequence 2, Appl1
40	193.5	4.0	749	2	US-08-738-975-2	Sequence 2, Appl1
41	193.5	4.0	749	2	US-08-728-626-2	Sequence 2, Appl1
42	193.5	4.0	749	3	US-08-808-599A-2	Sequence 2, Appl1
43	181.5	3.8	1248	2	US-08-348-153-17	Sequence 17, Appl1
44	181.5	3.8	1248	2	US-08-465-965-17	Sequence 17, Appl1
45	181.5	3.8	1248	3	US-08-465-966-17	Sequence 17, Appl1

## ALIGNMENTS

RESULT 1  
US-08-728-470-9  
Sequence 9, Application US/08728470  
Patent No. 5928651  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sioemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,470  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0810  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1338 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-728-470-9

**BEST AVAILABLE COPY**

---

```

Db 361 SSSSSSSVSEAPLATSSVSEAPSSSTSSSVSEAPSSSTSSSVSEAPSSSTSSSVSEI 420
Qy 277 TLTONOML-----LFNNNT-----STAGGAIYKKELASGGITLFSRNSVNGT 322
Db 421 SSTSSVSWSESVSSATSSLSSEAPSAISSLASRSLFSKNTSVTLVATEASSVSSL 480
Qy 323 APKGAIAIEDSGELSLADSGDIY-----FLGNTV-----TSTPCTN 361
Db 481 RPSEETLASNITSSLSSTGNTSVSTTSAASSTLSKSSNSKRAKTSSTSSDLS 540
Qy 362 RSSIDLTSAKMTALRSA-----AGRAIYFDPITGSSSTTVTLVKVNETPAD-SALQY 415
Db 541 KSVYFGNSSTVTSPSASISLTFASPLPSVMSDITSSPSSISSNLASSAPSDNNSTIA 600
Qy 416 TGNITFGEKLTSEADSKULSKLQPVTLSCGTLISLKGVTLLQQAFLQADSRLEM 475
Db 601 SASLIYTKTKNSVSSIVSSITSETNESMLATSSSTLSNKAATARSLSNATASNV 660
Qy 476 DVGT-----TLEPADTSTINNVLIN-----ISSIDGAKKAKIETKATSKNLTLIS--- 519
Db 661 PTGTFSSKMSHTVITTEGFTSSASLAINSTVSSSLAGISFTSPESPTTSTLVTSAP 720
Qy 520 GTITLDPPTGTFYENHSLRNP-----QSYDILELKASGVYVST 557
Db 721 STVSMSTTSAPFINNSTARPSPSTASPTSTSTSSISVPL-ASGDVYSS 769

```

```

RESULT 15
CBPA_CLOCL
ID CBPA_CLOCL STANDARD; PRT; 1848 AA.
AC P38058;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
GN CELULOSE BINDING PROTEIN A PRECURSOR.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium
OX NCBI_TaxID=1493;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92228810; PubMed=1565642;
RA Shoseyov O., Takagi M., Goldstein M.A., Dol R.H.;
RT "Primary sequence analysis of Clostridium cellulovorans cellulose
RL binding protein A."
RL Proc. Natl. Acad. Sci. U.S.A. 89:3483-3487(1992).
CC - FUNCTION: BINDS TO CELULOSE FIBERS AND COORDINATES CELLULOSE
CC ENZYMES.
CC - P-TM: THE N-TERMINUS IS BLOCKED.
CC - P-TM: CONTAINS ABOUT 10% OF CARBOHYDRATES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M73817; AAA23218.1; -.
CC PIR: A44140; A44140.
CC DR HSSP: Q06851; INBC.
CC DR InterPro: IPR001956; -.
CC DR InterPro: IPR002102; -.
CC DR Pfam: PF00942; CBD.3; 1.
CC DR Pfam: PF00963; Cohesin; 9.
CC KM Cellulose degradation; Cell wall; Repeat; Signal; Glycoprotein.
CC FT SIGNAL 1 28
CC FT CHAIN 29 1848
CC FT DOMAIN 29 189
CC SO SEQUENCE 1848 AA; 189152 MW; 85FA6CE6F771AFLA CRC64;

```

```

Query Match 4.5%; Score 215; DB 1; Length 1848;
Best Local Similarity 20.7%; Pred. No. 0.00031;
Matches 167; Conservative 84; Mismatches 262; Indels 294; Gaps 33;

Qy 21 IATEVLDSSASFDGNKNGNFSVRESQEDACTYLLFGKNTVLENIPTGTATKSCFENNT 80
Db 193 VPSSINPTSAFPKNTV-----KQADVKTMTLNGNTEKTKITDANGTALNASTDYSV 245
Qy 81 KG-DLFT-----GNGNSLFFQVD-----AGTVAGAAVNSS- 111
Db 246 SGNDVTISKAVLAKOSVGTTLNFNFSAGNPQKLVTVVDPEVAVATIGKQVGNAGET 305
Qy 112 -VNDKSTFFIGFSSLSFLASPGSSITTGKAGVSCSTG-----SLKPEKXNS-----LLF 159
Db 306 VAVPNLTKVPAAGLATELPLTFDSASLEVASITAGDIYNLPSVNSSTVSGSTIKLF 365
Qy 160 -----SKNFTDNG--GATKATKLSLGTGTMAL----- 186
Db 366 LDDPLGSQLTKDGVFAITTFKAKAITGTAKVSVKLATGTPVVGDAQLQEKPCAVNPGT 425
Qy 187 -----FSENTSKKG-----ALQTSDA----- 204
Db 426 VTINPIDNRMQISVGTATVKAGETAAVPTLTVSPSTGIATAEAOVSFADATLLEVASYA 485
Qy 205 -----LITGNGGEVSFSDNNTSSDGAIFTEASVTLSNNAKVSFIDNKYT-- 250
Db 486 GDIYLNPTVNFSTYVNGVNIKLFDDT---LGSQLSKDGVEVYINFKAAVSTVYTP 542
Qy 251 -----GASSSTTGDMSGAICA-----YKSTDTKRVTLTG 280
Db 543 VTSGTPEAFADGTLAEVQSKNAAGSVINIGDPILEPTISPVTATPFKAKADAVATMTL 602
Qy 281 NQMLFSSNNTST-----AGGAIYKKELAS---GGTLFSRNSVNG- 321
Db 603 NG-YTFNGITGLTSDYSISGNVVKISQAVLAKQPVGDITL-TFNFSNGKNTATAKLVVS 660
Qy 322 --TAPK-----GGAIAIE-----DSGELSLSD-----SGDIY 347
Db 661 IKDAPKTYTATVGTATVAGETVAVPYTLNSVSGISTAEQLQSLFADATLLEVASTAGDIY 720
Qy 348 FLGNTVITSTPGTNRSSIDLTSAKMTALRSAGRAIFYDPI-----TSGSSTVYTD 400
Db 721 L-----NPSVFSVSVNGSTIKLFLDPTLGSQLSKNGVATINFKAKSVSTVYTT 772
Qy 401 VLKVNETP--ADSAIQYTGNIITFGEKLTSEADSKNLSKLLQP---VTLSCGTLISL 455
Db 773 PVKVSCTPEVFADGTL-----AELSYETVAG---SVTINAIGPVKTVATVGTATVK 820
Qy 456 HGVTLQQAFTQO-----ADSRLEMDVGTLEPADTSTINNVLINIS----- 497
Db 821 SGETVAVPVTLNSVPGIATAEQLQSLFADATLLEVASTAGDIYLPVNSVSSVNGSTIKL 880
Qy 498 -----SIDGAK-----KAKIETKATSKNLTLSCGTLTLLDPGTFYENHSLRNP 540
Db 881 LFLDDTLGSQLSKDVGLATINFKAKVTSVYTPPAVSGTPVAD--GTLAEQLQSTVA 938
Qy 541 QSYDILELKASGVYTAVPDPIMGE 567
Db 939 GSVYIEPSQPVKTVATVGTATVKSGE 965

```

Search completed: October 2, 2001, 03:32:39  
 Job time: 274 sec

```

FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1041 EGT2 PROTEIN.
FT DOMAIN 200 203 POLY-SER.
FT DOMAIN 381 384 POLY-SER.
FT DOMAIN 388 395 POLY-SER.
FT DOMAIN 490 493 POLY-SER.
FT DOMAIN 586 589 POLY-SER.
FT DOMAIN 457 492 POLY-SER.
FT REPEAT 577 606 1-2.
FT REPEAT 613 647 1-3.
FT REPEAT 716 745 1-4.
FT REPEAT 773 802 1-5.
FT REPEAT 811 840 1-6.
FT REPEAT 849 886 1-7.
FT REPEAT 887 924 1-8.
FT REPEAT 925 962 1-9.
FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 526 526 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 544 544 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 636 636 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 657 657 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 709 709 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 1041 AA; 108494 MW; 01FECEFB8A8744CD CRC64;

```

Query Match 4.5%; Score 216; DB 1; Length 1041;  
 Best Local Similarity 23.0%; Pred. No. 0.00012;  
 Matches 164; Conservative 97; Mismatches 272; Indels 180; Gaps 32;

```

QY 4 SPKPFYFPAFLPLSMITER-----VDSASFPKNGNFVSQSE 48
DB 395 SAPSSNSFTTPSSSLSTATETYSISASISVTOASYDNSTTAVTOSTSIHVASAE 454
QY 49 DAGTYLFKGNLTLENIPGTGAITKSCFNMTKGLTFGNGSLIFQYVDA-----GTV 103
DB 455 KUSSTLSTYISVNTISVSATQHTTTPSYVNST-----TLSSSVLESVISSPYLANTTV 509
QY 104 AGA-----AVNSVVDKSTTFIGFSSLSFIAS-----PGSSLT---TGKAVSCSTG 147
DB 510 SGASSASQSTNPYPYVNST-----SSATQLATIAPAINITGSISSITNNSSVSSTIS 565
QY 148 SLKFDKNSVLLSKNFSTNGCAITAKTSLTGTTMSALFSENTSKKGAIGTSDALTI 207
DB 566 SLSSGPFVS-----NTAVASGSYLLTFTTESAOLTE-----ISLIPISITTS 609
QY 208 TGNQGVSVSDNTSSDGAIFTEASVTISNNAKYS---FIDNKVYGAASS-----TTGD 259
DB 610 TTTSG-----TDITGSNK-VASSTELIAGSYVNSLSVSTINNTAATAANARATTTAT 664
QY 260 MSGGALCAIKYST-----DFTVTLTGNOMLFSNNTSTAGGAIYVKLELASGGLTLE 313
DB 665 HSGSLQPSYHSSLSLSTIDTKYT-----TATTSRSDGS-----SLAFTTG 707
QY 314 SRNSVNGGAPRGGAIAIDSGELISADSGDIVLGNTVTSTPEOTNR-SSIDLGTSK 372
DB 708 LMSVYTGK-DKSDTYSVISTESAGVTEYDILPLISTLKPIYVGTCTSHNSTFSVWSYK 766
QY 373 MT-ALSSAGRAIFYDPTTGSSTTVTD-----VLKVE-----T 407
DB 767 LPEATITDGDGA---YSVISTQSAGVTEYGMPLISTLETPIVIMSTDESGYILTCT 823

```

```

QY 408 PADSALOY-----TGNII--FTGEKL-----SETEADSKNLTSKLLQPTLSGTL 452
DB 824 ESGQAETVSLPISTLSDSVITFTGESVVGYSITVGAQAQHTSLVSPVSTIKSGT 883
QY 453 SLKHGVTLOTQAFQOQADRLMDVGTTLPEPADTSTINNLVINISSIDAKAKAKIETKAT 512
DB 884 SL-----STESVYAGYSTIVGAQAQHTSLV-PVSTIKSGK-----TSLIS 924
QY 513 SKNLTLSTITLLDPTGTFYENHSLRNPOSYDILELKA--SGVTSTAVTAPPIMCKPH 570
DB 925 TEESVYAGISTVVD--SAQYAEHT--NLVAIDTLKSTQKATFATVCTCARLSSPSPA 980
QY 571 YGYOGTGPVWGATGASTATATFNMTKGYIPME-----RIGSLV 610
DB 981 TLDAGTITSLPTSSSTSLSTIITWYSSSTIKKPSISTYSGAAGQLTIRIGSL 1033

```

```

RESULT 14
Y96_YEAST STANDARD; PRT; 1140 AA.
ID Y96_YEAST
AC 004893:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 113.1 KDA PROTEIN IN PRB5-FET4 INTERGENIC REGION.
GN YMR317W OR YMR924.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
ON NCBI_TaxID=4932;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972.
RA Churcher C.M., Louis E.J., Barrell B.G., Rajadream M.A., Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
CC -1- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
DB EMBL; 254141; CAA90835.1; -.
DR SGD; S0004936; YMR317W.
KW Hypothetical protein; Repeat.
SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

```

Query Match 4.5%; Score 215.5; DB 1; Length 1140;  
 Best Local Similarity 20.9%; Pred. No. 0.00015;  
 Matches 136; Conservative 110; Mismatches 285; Indels 119; Gaps 20;

```

QY 18 LEMITET-----VDSASAFDGNKNGNFVSRESQDAGTYLFGKNTLENIPGTGA 71
DB 129 LSSSTETDISISSAIGTSSPQTSSSNGGS--SSEPLGKSSVLETTASSSDT---TA 181
QY 72 TRKSCFNMTKGLTTFGNGNSLIFQTVADGYAGAA---VNSVVDKSTTF----- 119
DB 182 VYSSTF-TLTPVSSSPKISSGSAVTSYGTTSDAKKEYFSSSTSPVSSLSSTSSPASS 240
QY 120 -----IGFSS--LSFIASP-----GSSITPGKAVSCSTGSLKFDKNVSLFSKNE 163
DB 241 TISETLIPRSTLSTLSTSSVSSEAPSPATSSVSSEASSSTSSVSSEAPLAPLSSVSSA 300
QY 164 STDNGCAITAKTSLTGTTMSALFSENTSSKKG-----GAIGTSDALITGNGGVSE 216
DB 301 PSTTSVSVSEAPSPSTSSVSSEISSTTSVSSSEAPLATSSVSEAPSPSTSSVSSEI 360
QY 217 SNTSSDGAALFTEASVTISNNAKYSFIDNKVYTGASSSTTGDMSGGALCAIKYTIIDTV 276

```

RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,  
RA Sempel G., Seki Y., Sivasubramanian S., Tagami H., Takeda J.,  
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.,  
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 28.0-40.1 min region on the linkage map.",  
RL DNA Res. 3:363-377(1996).  
RP [3]  
RP SEQUENCE OF 464-2003 FROM N.A.  
RC STRAIN-K12:  
RX MEDLINE=92190338; PubMed=1665988;  
RA Moszer I., Glaser P., Danchin A.;  
RT Multiple insertion sequences near the replication terminus in  
RT Escherichia coli K-12.";  
RL Biochimie 73:1361-1374(1991).  
CC -1- SIMILARITY: TO S. TYPHIMURUM ORF NEAR CYS6 (AC P25928).  
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR  
CC THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS3C ELEMENT  
CC BETWEEN AMINO ACIDS 839 AND 840.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AE000337; AAC74483.1; ALT\_SEQ.  
CC EMBL: AE000337; AAC74487.1; ALT\_SEQ.  
CC EMBL: D90778; BAA15009.1; ALT\_SEQ.  
CC EMBL: D90778; BAA18880.1; ALT\_SEQ.  
CC EMBL: D90779; BAA18881.1; ALT\_SEQ.  
CC EMBL: X62680; NOT\_ANNOTATED\_CDS.  
CC Ecogene: EG11307; ydda.  
CC Hypothetical protein.  
CC FT CONFLICT 489 489 I -> V (IN REF. 2).  
CC FT CONFLICT 495 495 I -> V (IN REF. 2).  
CC SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;  
  
Query Match 4.58; Score 216.5; DB 1; Length 2003;  
Best Local Similarity 21.3%; Pred. No. 0.00029;  
Matches 165; Conservative 92; Mismatches 276; Indels 243; Gaps 35;  
  
QY 23 TENVLDSSA---SPDGK-----NGNSVRESQDACTYLFKGNVLEINPG 67  
DB 280 TMTVTDPESMGIOIDGKAIYNNEGESTITNGTGTQINGDA--TANNNGKTTVDGKDS 337  
QY 68 TGAATKSCFNMTK---GDLTFGNGNSLLFOTVDAGTVAGAAVSSVYDKSTTFIGF 122  
DB 338 TGTFTING--NNGKVIDDGDLDVSGGGHGDID--TGD SATVDNKG--TMTVTDPES--IGIQ 390  
QY 123 --SSLSFIASPGSSITTKGKAVSGSTGSLKFDKNVSLFSKFNSTNDGCAITFAKTLISLTG 180  
DB 391 VDGQDAVAVNNEGESAITNGTGTQINGD-----DATANNNGKTTVDGKSTG 437  
QY 181 TTMSALESEMTSSKKGAIQTS-----ALFTIG--NCGEVSPSDNTSSSGA 226  
DB 438 TET-----AGNNGVVIDDGDLDVSGGGHGDIDTGD SATVDNKG--TMTVTDPES---I 485  
QY 227 ALTFEASVTISNNKAVSFINDKVTGA---SSSTTGMSGCAICAKTSPDKKVTILTNOM 283  
DB 486 GIDIDGQALVNNEGESTITNGTGTQINGDANFANNNGKTTVDGKSTGKTI--AGNIG 543  
QY 284 LIFSNNSTTAGAIIYVKL-----ELASG--GLTLFSRNSVNGG 321  
DB 544 IYVLDGSLFTVGAHGVENIGDNGTNNKGDIVYSDTSGISGLINGGARVSNMGDVNS 603  
QY 322 TARKGGAIALEDGSELSD-----SGDIVFLGNNVTSTTPG 359  
DB 604 NEATGFSIT--TNSGKSLASMOVDEFTGVDLNNGNNSVTLAKDKLVVGGQKRTGINS 662  
QY 360 TNRSSIDLGTSAKMTALRSAAGRAIFYDP-----ITTSSTVTVDLAKVNET 407

DB 663 GDANTVNTGVLVLDKOKTADNAAEYFDPESVGINVGSNNVTLDGKLTFTVSDSEVTSR 722  
QY 408 PA---DSALQYTTGNIITFG-----EKLETERA-ADSKULTS-----KLLQPTL 447  
DB 723 QSNLFDSAEKTSGLVIYDGNITVNNNGLELLIGEKNALDSGCVTSIRGYSYTSYIV 782  
QY 448 SGGTILSLKHGVTLTQOAF-----TQOADSRLKEMDVTGLEPDPSTINNLTINISSI 499  
DB 783 SGESSVYINGTTISGERPLGFAGYIRYQKALLFISGATLTQMODIDSEH----- 834  
QY 500 DGAKKAKIETK---ATSKNLT--LSCGTTLL---DP-----T 528  
DB 835 HGTREVEIONGFAPVNGENTGTNGSITSLONGKDPAPSPYVLLATNGGSAATNAGTIT 894  
QY 529 GTFENHSLSRPOSDIIELEKASGTIVSTANTPPPIGKEKHYGQGWGIVGTCAST 588  
DB 895 GKVTQHSVFENKYS-----TGTNSNFIFFNDV-----SSITGLVAQS 931  
QY 589 TATENMTKRTGYI-----PNERIGSLVNSLW-----NAFIDISS 623  
DB 932 NSTLINDSGIIDLKGSVGMALADSTAEKCKITLDSKMDVANDTTAMRDIAS 987  
  
RESULT 13  
ID EGT2\_YEAST STANDARD; PRT; 1041 AA.  
AC P42835;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE EGT2 PROTEIN PRECURSOR (EARLY G1 TRANSCRIPT 2).  
GN EGT2 OR YNL327W OR N0320.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / FY1676;  
RX MEDLINE=95373280; PubMed=7645347;  
RA Mafthil M., Nicoud J.-M., Levesque H., Galliardin C.;  
RT "Sequencing analysis of a 15.4 kb fragment of yeast chromosome XIV  
RT identifies the RPD3, PAS8 and KRE1 loci, five new open reading  
RT frames.";  
RL Yeast 11:567-572(1995).  
RN [2]  
RP SEQUENCE FROM N.A. AND FUNCTION.  
RX MEDLINE=96251274; PubMed=8668141;  
RA Kovacech B., Nasmyth K., Schuster T.;  
RT "EGT2 gene transcription is induced predominantly by Swi5 in early  
RT G1.";  
RL Mol. Cell. Biol. 16:3264-3274(1996).  
CC -1- FUNCTION: SEEMS TO BE INVOLVED IN THE CORRECT TIMING OF CELL  
CC SEPARATION AFTER CYTOKINESIS. AS SEPARATION OF MUTANT DAUGHTER  
CC CELLS IS DELAYED. COULD EITHER BE AN ENZYME NECESSARY FOR GLUCAN-  
CC DEGRADATION OF THE CELL WALL AT THE NECK REGION BETWEEN MOTHER AND  
CC DAUGHTER CELLS OR A REGULATORY PROTEIN CONTROLLING THIS METABOLIC  
CC STEP. EXCLUSIVELY EXPRESSED BETWEEN THE END OF MITOSIS AND EARLY  
CC G1; INACTIVATED BEFORE CELLS PASS START.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: Z46259; CAAB6371.1; -;  
CC EMBL: Z71603; CAAG6259.1; -;  
CC SGD: S0005271; EGT2.  
CC Glycoprotein; Repeat; signal; cell cycle; cell division; Mitosis.  
KW

```

QY 722 PCKL--RPCSLSVPEHLEFSGNLSTHTDNDLTKR 757
DB 1500 PNVLYTPMAGLSY-----LKSSNENYKGTGTAYANK 1530

RESULT 11
FIG2_YEAST
ID FIG2_YEAST STANDARD; PRT: 1609 AA.
AC P25653:
DT 01-MAY-1992 (rel. 22, Created)
DT 01-MAY-1992 (rel. 22, Last sequence update)
DE 15-DEC-1998 (rel. 37, Last annotation update)
DE FACTOR INDUCED GENE 2.
GN FIG2 OR YCR089M OR YCR89M OR YCR1102.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92397594; PubMed=1523889;
RA Wilson C., Grisanti P., Frontali L.;
RT "The complete sequence of a 6146 bp fragment of Saccharomyces
RT cerevisiae chromosome III contains two new open reading frames.";
RL Yeast 8:569-575(1992).
CC -1- FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC -1- INDUCTION: BY MATING PHEROMONES.
CC -----
CC This SMSS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X59720; CAA42254.1; .
DR PIR: S19504; S19504.
DR PIR: S25345; S25345.
DR SGD: S0000685; FIG2.
SQ SEQUENCE 1609 AA; 166049 MW; 7D66AD7F85A7B852 CRC64;

Query Match 4.7%; Score 222.5; DB 1; Length 1609;
Best Local Similarity 21.1%; Pred. No. 9.9e-05;
Matches 206; Conservative 128; Mismatches 365; Indels 279; Gaps 43;

QY 11 STPAIFPLSMI---ATEVLDSSASFQKNGNFVRESQEDAGTTLFKGNVLTENING 67
DB 362 AATSLIDPISVSGSTASSFVGISTANFSTGNSNY-VPESTAGSSQYQDMSSSLPLSQT 420
QY 66 TGTATKSCFNNTKGLPTTNGN--SLLEFQVADAGTGAANNVSVYKSTTFIGFSSL 125
DB 421 TWWVINTT---NTQGSVSTSPAYVSTAFKTVD-GVITEYVWCPPLQTKSOAIGVSS- 475
QY 126 SFIASPGSSITTKGKAVSCSTSLFKDKNV-----SLLFSKNF----- 163
DB 476 STSSVPOASFSFGSSSTLSSNSTLAASNNVPESTAGSSQYQDMSSSLPLSQTWWVIN 535
QY 164 STDNGCAITAKT-----LSLGTGYMALESENT 191
DB 536 TINTQGSVSTSPAYVSTAFKTVDGVITEYVWCPPLQTKSOAIGISSSTISATQTSKP 595
QY 192 SS--KKG-GAITQDMLITTGNO-----GEV----- 214
DB 596 SLLILGISTLQISDA-TFKGTETINTHLMTESTITEPTTSGISDSYLYLTSEVNLAS 654
QY 215 ---SFSQNTSSDGAIFTEASVTISNNAK-----VSFLDNKYTGASST 256
DB 655 SLSSYNFSSSEGSTATINTSVTGSTSKYPTSVNPTFASQHVSSSVNSLTDFSTNS 714
QY 257 TCGMSGALCA--YKTS-----DTFKVTLTGNOMLFSSNNTSTTACGA----- 297

```

```

DB 715 TEETI---AVISNIHKTSNKDYSLATTQDKTSGKQTLVLSVTYVNGAATEYTWCPAS 771
QY 298 -----IYVKLELASGCL-----TLFSRNSVNGTAPKGAIAIDSGELSLADS 343
DB 772 STAYTTSISKVLVLTTEVCSHSECTPIYITSVTATSTIP---LISTSSVLSYSE 828
QY 344 G-----DIVFLGNVTSTTPGTRNRSSIDLGTSAKNTALRASAGRAIYEDPITGSSST 397
DB 829 GAKNPASAEVITINTQVSAISEATSTSTQVSAATATASES-----TTSQVST 877
QY 398 VVDVLKVNFT-----PADSALOYTGNIITFGEKLSETEAADSNNLSKLLQPT 446
DB 878 ASETITLTQONFTTYSLLFPALSTEMINTVVSRTLLISTEVCSHSCVPTVITEV 937
QY 447 LSGGTLSLKH-CVTLQTOAFQOADSRLMDYGTTLLEPADTSTINLVY----- 494
DB 938 TSKGTPSNHSSQTLQTEAVEVTLSSHQTVMTST--EVCNSICTPTVITTSQMRSTPP 995
QY 495 -----NISSIDGAKKAKIETKATSKNLTLSGTTLLDPTGTFEYENHSLRNPQSD--IL 546
DB 996 YLTSSSTSSSLASTKKSLE--ASSEMSFVSSTQSLPLATFCSEKRSYTSVSQMENTVL 1053
QY 547 E---LKASGTVSTAVTPPIGME-KFHGYQ--GTMGPIWNGTGASTTATFMTGTG- 599
DB 1054 TWTIMSSSNVISTNEKPPSTSPYNFSSGYSLPSSSTPSQYSLSTAT--TINGITVYT 1112
QY 600 --IPNER-----IGSLVNSLMAAFIDISSLHYLMETAN---EGLGDRAFW 642
DB 1113 TWCPLAEKTVAASSQSSSVDRFVSSSKPSSLSQFSLQYLTSLTFTTISGLKTYTTW 1172
QY 643 CAGLSNFFHKDSTKTRGRFRLHSGGYVIGNLHTCS--KILSAFC-----QLF 690
DB 1173 CP-----LTSKSTLGATQTSSTAKVRLTSSASASTSISLSTSE 1213
QY 691 GADRDYFVAKNGTYTGGLLYQHNETYISLQCKLRPCLSLVPTPIPLFSGNLSY-TH 749
DB 1214 SESSSGYLSKG---VCSGTCEQDVPQTQSSPAS---FLAVSPS---VSTSSSSFSFT 1263
QY 750 TDNDLTKRTYTPYTVKGS 767
DB 1264 TASTLTSTHTVPLPSS 1281

RESULT 12
YBBA_ECOLI STANDARD; PRT: 2003 AA.
ID YBBA_ECOLI
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (rel. 28, Created)
DT 01-OCT-2000 (rel. 40, Last sequence update)
DT 01-OCT-2000 (rel. 40, Last annotation update)
DE HYPOTHETICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.
GN YBBA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley J.M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,

```



DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)  
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
GN OMPB OR SPAP OR SPA OR RP704.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiidae; Rickettsia.  
OX NCBI\_TaxID=782;  
RN [1]  
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
RC STRAIN=BREINL;  
RX MEDLINE=91045972; PubMed=2122457;  
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;  
RT "Characterization of the gene encoding the protective paracrystalline-  
RT surface-layer protein of Rickettsia prowazekii: presence of a  
RT truncated identical homolog in Rickettsia typhi.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BREINL;  
RA Moron C.G., Yu X.J., Walker D.H.;  
RT "Sequence analysis of ompb of Rickettsia prowazekii.";  
RL submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MADRID E;  
RX MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Slicher-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RT mitochondria.";  
RL Nature 396:133-140(1998).  
RN [4]  
RP PARTIAL SEQUENCE.  
RC STRAIN=BREINL;  
RX MEDLINE=92114896; PubMed=1370573;  
RA Ching W.M., Carl M., Dasch G.A.;  
RT "Mapping of monoclonal antibody binding sites on CNBr fragments of  
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
RT prowazekii.";  
RL Mol. Immunol. 29:95-105(1992).  
RN [5]  
RP IDENTIFICATION OF CLEAVAGE SITE.  
RX MEDLINE=92104668; PubMed=1729180;  
RA Heckstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;  
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer  
RT membrane protein of rickettsiae: identification of an avirulent mutant  
RT deficient in processing.";  
RL Infect. Immun. 60:159-165(1992).  
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
CC VIRULANCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
CC LAYER WITH HEXAGONAL SYMMETRY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M37647; AAA26390.1; ALT\_INIT.  
DR EMBL: AF161079; AA42234.1; -  
DR EMBL: AJ235273; CAAL5140.1; -  
KW Antigen; S-layer; Cell wall.  
FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.

FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.  
FT VARIANT 257 257 V -> A (IN STRAIN BREINL).  
FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).  
FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).  
FT CONFLICT 178 179 AA -> VC (IN REF. 1).  
FT CONFLICT 191 201 TTGCAPLTGA -> INSRSSYLVS (IN REF. 1).  
FT CONFLICT 212 212 T -> I (IN REF. 1).  
FT CONFLICT 313 313 O -> L (IN REF. 1).  
FT CONFLICT 1104 1104 D -> G (IN REF. 2).  
FT CONFLICT 1123 1123 T -> S (IN REF. 2).  
SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;

Query Match 4.7% Score 223; DB 1; Length 1643;

Best local similarity 20.7%; Pred. no. 9.6e-05; Matches 182; Conservative 131; Mismatches 327; Indels 238; Gaps 44;

QY 23 TETVLDSSASFDGKMKNGFVSRESQEDAGTYLFGKGVY----- 61  
DB 748 TSIYSGTVGGQGHKLNLL-----DNGTVKFLGDTTFNGGTRKECKSLIQISNNVYT 802  
QY 62 --LENIPGTG-----AITSCKFNNTKGLTF--GN----- 89  
DB 803 DHVESADNTGTLFEVNDPITVTLNKOGAYGVLYISGPGNIVFENIGNVHICIA 862  
QY 90 GNSLFFQTVDA-----TVAGAAVNSVVDKSTFTLIGSSLSFIASPGSSI 135  
DB 863 ANSISFENASLGTSLEPSPGLDVLITKSTVGNGTVDNFAPVYVVSIGDISMINNGIT 922  
QY 136 TTGGAVSCSTGSLKFDKNVSLFSKFNSTDNGAIFAK-----TLSLTG-----TTMSA 185  
DB 923 GDKNNIIALSLGS---DSISIV---NANTLYSGIRTRKNNQGVITLSCGPNPNPGITYG 975  
QY 186 LFSENTSSKKGAIQTSDALITTNQO-----EVSFSDNTSSDGAALFT--EASVT--- 235  
DB 976 LGLENGSEKRLQVTFETD---YNNLGSIIANNVTINDYVTLTGAGTDFDAKITLGS 1031  
QY 236 ISNAKVSFIDNKVTG-----ASSSTGDMSS--GGAICAKKTSNDT--KVTLGNQML 284  
DB 1032 VNGNANNAFVDSTFSDPSRMVATQANKGVITYLGNALVNSIGSLDPVAVSRVTG----- 1087  
QY 285 LFSNNTSTTAGAIVYKVLKELASGLTLFNSNSVNGTAPKGAIAIEDSGELSLASDGS 344  
DB 1088 ---NDSGAGLGQGNISQMDHGFYNLITLNSNVILG-----GGTTAI--NEIDILL--TTN 1135  
QY 345 DIVFLNGTVITSTTPTGNSSIDLGTSAKMTALRSAGR--AIYYDPTTSSSTTVYDL 402  
DB 1136 NLIFANGT---STWGDNTSI-----STTLNVSSGNIQGVIAEDPAQVNAVATYGTITI 1184  
QY 403 KVNE-----TPADSALQ-----YGNIFTEGKLESETPAAD--SKMLT 438  
DB 1185 KIQNNANANFSGTQATYTLIOGGAFFNGLGNPFAVYGSNIFVYELIRSDNOYVLTRT 1244  
QY 439 SKLLQPTVTLSSGTLISLKH--GVLQLOQAFTQOADSRLMDYGTLEPDAVSTINNLYI-- 494  
DB 1245 NDVLNVTTAVAGNSAIANAPGVS-----QNSRCLSESTNPAAYNNMLLAK 1289  
QY 495 ---NISSIDGAKKAKIEFKATSKNL--TLSGTITLDPGTGFENHSLRNQSDILELK 549  
DB 1290 DPSVAIFVGAIAITDTSAAVTVNLNDPTOKTODLLSNRLGTL---RLLSAETSDVAG-S 1345  
QY 550 ASGTVSTAVVPPDPMGEKFFHYGQTMGPVWGTAST-----TATFNWTKYGIIPPER 605  
DB 1346 ATGAVSS-----GDEAEVST--GVNAKFFYIAEDOKKGGTAGAKATTGVVGLDT 1395  
QY 606 IGSIVPSLSLWNAFIDISLHAYMETANEGLOGDRAFCAGLSNFFHDKSTRTKRGFRHLS 665  
DB 1396 LAS--DNLMIGALIGITKTDIKHODYKKGDKTD-----INGLSFELY-----GSQGLV 1441  
QY 666 GGYVIGNLHTCSKILSAA---FCQLFGRDRDYFAKN--OGTVYGGTLYYQHNEYIYSL 721  
DB 1442 KNFPAQGNALFTLNKRVKSKSQRYFFESNGKSKOIAAGNYDNMTFNGNLIFGQYD--YNAM 1499

```

DB 768 ---DNGTTFVKELGNATFENGNTTIANSTLOISGNTADFIASAGTGIVE--VWTG-- 819
OY 396 TTYTDLKAKVENPADAALOYT---GNITFEDEKLESEPAASKNLTSLLQPVTLIS--- 448
DB 820 -PINTVLNKQAVNVAVKOITVSGPQNVV-----NEIGNAGNTHGAMTDTIAFENSS 871
OY 449 -GGTSLKHKVTLQTOAFTQADSRLEMDVGTLEPADSTINNLV-----INISSI 499
DB 872 LGAVFLPBGIFPN-----DAGNTIPLTIKSTVGNTEBEGFSVPVIVSGV 917
OY 500 D-----GAKKAKIETKATSKNLITSGTITLDPGTETENHSLNPQSYDILELKA 550
DB 918 DSVIADGQVIGQNNIVGIGLCSGDNCIYVATTLAIGITINN----- 961
OY 551 SGTIVSTATPDPINGEKHYGQWGPVWGTGATATFNMKTGYIIPERIGSLV 610
DB 962 QGTIVLGGVPT-----PGTVYGLGTGIGASKFKQVFT-TDY-----NNIGIT 1006
OY 611 PMSLMNAFIDISLHYLMETANGLQ---GDRAFWCAGL-----SNFPHK 652
DB 1007 ATN-----TTINDGYVTGTGIAAGIAGTDDGKTITLGSVGNANVRFA 1051
OY 653 D-----STKTRRG-FRHLGGVYIGNLHTCSDKILSAFCQLFGDRDIFYAK 700
DB 1052 DGIFNSTSMITVTKRANGTVYLGNAFV--GNIGSDPVPVASVR-----FTGS 1098
OY 701 NCGTVYGGTLYVOHNE--TYISLPCKLRPCSLSYVTELPVLEFSGULSTHTDNLKTYK 758
DB 1099 NNGAGIKGKIYQVDEGIVY-----NLGIYNSN--VILGGSTTAINGKIDILLNT 1146
OY 759 TTYPTVKGSGND-----SFALFEGGRAPICDESA 789
DB 1147 LTFAGGTSTWGNNTSLETTLLTLANGNIGHIYIAEGA 1182

RESULT 9
ID Y109_YEAST STANDARD; PRT; 995 AA.
AC P40442:
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHEICAL 99.7 KDA PROTEIN IN SDL 5' REGION PRECURSOR.
GN Y1169C OR Y19402.07C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Genies S., Hamlyn N., Horsnell T.S., Hunt S., Jagsis K., Jones M.,
RA Louis E., Lye G., Moule T., Moule T., Odeil C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL; 246921; CAA87023.1;
DR SGD; S0001431; Y1169C.
KW Hypothetical protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 995 POTENTIAL.
FT HYPOTHETICAL PROTEIN Y1169C.

```

```

FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 468 468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 664 664 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 995 AA: 99735 MW: F63E287A03F137EC CRC64;

Query Match 4.7%; Score 226.5; DB 1; Length 995;
Best Local Similarity 21.1%; Pred. No. 3,1e-05;
Matches 173; Conservative 143; Mismatches 343; Indels 159; Gaps 35;

OY 29 SCASFDGKNGKNGFVSRESDAGTTLFEGNVTLENIPEGTGAIKSCNNKGLDITFG 88
DB 37 SSTSVSSSSSGSVSSISSIAE-----TSSATDILSIITQSA--STSGVSSSVG 84
OY 89 -NGNSILPQTVAGYAGAAVNSVYDKSTFTLGESSLFISAPGSIITGGAAVCSNG 147
DB 85 PSSSVSSSVSSQSSSVSSVSSQSSSASDVS--SSVQASASTISDVSSSVSSQSS 142
OY 148 SLKPKNNVSLFSKNEFTDNGCAITAKTLTGTWMSALESENTSKKGAIOQSDALTI 207
DB 143 SAS-DVSSSVSSQSSSASDVSSSVSSQSSASASDVSSSVSSQSSASTIDVSSSVSSQSS -SA 200
OY 208 TGNQGEVFSFSDNTSDSGAIFTEASVTISNNAKVSEIDNKVTGASSSTTGMS--GGAI 265
DB 201 SPVSSSVSSQSSSASDVSSSVSSQSSASTIDVSSSVSSQSSASTIDVSSSVSSQSS 260
OY 266 CAYKSTQDKVLTG-----NOMLLPNNNTSTAGAIYKKEIASGGLT 311
DB 261 SSFPOSTSSASTASGANSLSITSSASASAFASNSLSSSDGTYLPTPTT--SGDLT 319
OY 312 LFSRNSVNGTAPKGA-IAIEDSGELSLAD---SGDIVFAGNTVSTTPT--NRSSI 365
DB 320 LKGVIATGCVVVAAGAKITLLDGDKYFSADLKTYGDL--LVKSKKEYIPEETPDISG 377
OY 366 DVGTSKMTALPSAAGRA-TIYYDPTTSSSTTVYDVL-----KVENPADAALOYT-G 417
DB 378 NFDVYGNFNAEESAAITSASISYFTSPSPNSDISLSLSKSKGVEYTFPYSNCGAFSFS 437
OY 418 NIIFTEKLESEPADSK-----NL-----TSKLQPVTLG-----GTLILKHG 458
DB 438 NAILNGSVSGIQRDRDEGSVNNNGEINLNDGSTIYIYVPSGKTGVNLSNLVLIHYD 497
OY 459 TLQTAFTQADSRLEMDVGTLEPADSTINNLVINISSIDGAKKAKIETKATSKNLTL 518
DB 498 TETGGVYVKKGGVLAVD-----PTEI--NATPIPVGYGKQNIATADITA--LSY 546
OY 519 SGTITLLDPT-----GTFYENHSLNPQSYDILE-----LKASGYTSTA 558
DB 547 DGTGVLTAQGNRQPSFAIGTGFSS-----SDFSVSGEIPAGAVAYVLANGVATSA 600
OY 559 VTPDPIMGKPHYGQGTGPIWGTGASTTAPFNMKTGYIIPERIGSLV 618
DB 601 AASSSTRASGASAVTGSTSGASVYTGSTAST--SFGASVYTGSAIS--FASVYTG--TS 654
OY 619 IDISSLHYMETANGLQDRAFWCAGLSNPFHKDSTKTRGRFRLSGGYVIGNLHT-- 676
DB 655 VYTTLDIYVNAISVYVS-----CS-----ETDLS-----NGVYTTI 687
OY 677 ---CSDKILSAFCQLFGDRDIFYAKNGTVYGGTLYVOHNETYISLPCKLRPCSLSY 732
DB 688 TTPVPSSTTATITSCDENG---CHVSTGAVVEYVSSKSYTATVTHDCDNCNTKT 743
OY 733 VPTELPVLEFSG-----NLSTYH--TDNDLTKTYTT 760
DB 744 VTSSEKETSATTPASPKSTTYVTYVTHCDNGCNTKTYT 781

RESULT 10
OMP_RICPR STANDARD; PRT; 1643 AA.
ID OMP_RICPR
AC Q53020; Q9ZCNO;
DT 30-MAY-2000 (Rel. 39, Created)

```

DR EMBL: Y00544; -: NOT ANNOTATED\_CDS.  
 DR Ecogene: EG12850; yfal.  
 KM Hypothetical protein; Repeat; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 1250  
 FT DOMAIN 919 948  
 FT CONFLICT 28 30  
 FT CONFLICT 40 40  
 FT CONFLICT 65 66  
 FT CONFLICT 431 431  
 FT CONFLICT 433 434  
 FT CONFLICT 478 478  
 FT CONFLICT 773 773  
 FT CONFLICT 853 853  
 FT CONFLICT 923 924  
 FT CONFLICT 948 994  
 FT SEQUENCE 1250 AA; 131152 MW; 17F98C05E299FC95 CRC64;  
 LRVIRG -> LITSRC (IN REF. 4).  
 PATOPVIAKVGSLINLRANAKAFMEREDHAGGDSQTIN  
 Query Match 4.8%; Score 231; DB 1; Length 1250;  
 Best Local Similarity 24.1%; Pred. No. 2,4e-05;  
 Matches 134; Conservative 84; Mismatches 209; Indels 130; Gaps 28;  
 QY 102 TVAGAAVNSVVDKSTFIQSSLSFIAPSGSSITTKGAVSCSTGSLKFDKNVSLFSK 161  
 DB 49 SLGSIITDDMSIADQ--WLFPSDMTNAS--GGAVFLQGA--EFLSLPEHETGNTLFAN 102  
 QY 162 NFST--DNGCAITAK--TLSTLTGTMALFSENTSSKKGAIGTSDALTITGNGEVS 215  
 DB 103 NTVGEVNGCAIFAKENSTLNTL--DVIFSGNVAGYGAISGSG--TMDTGAVD 154  
 QY 216 -----FSNTSSD--SGAIFTEASVTISNNAKVS--FLDNKTVASSSTGSDMGGA 264  
 DB 155 LRYTNAMFRNNIADGGGAIY-----TINNDYLSYVIDNNQAYTSTISDGD--GGA 207  
 QY 265 ICAVKTSTDK---VLTGNOMLFSNNTSTTAGAIYVKLEL-----AS 307  
 DB 208 IDVTDNNSDKHPSGYITVNN--TAFNTNTEAGYGAIIYNSVAPYLLIDISVDYSQN 265  
 QY 308 GGLTFSRNSVNG---GTAKGGAIAIEDSGELSLADSGDIYFLGNTVSTTPGTRSS 364  
 DB 266 GGVLYDENNSAAGYDGPSSAAGFMYLGLSEVFYDIADKTLIVIGNT-----ENDGA 318  
 QY 365 IDGTSAKMTALSAAGRAIYFYDPIFTGSSITYVDLAKNENPADALQYTGNIITGE 424  
 DB 319 VD--STAGTGLTKTGS--DLVLNADNDPTGEMQIEN--GEVTLGRSNLNVGD 369  
 QY 425 KLSTEADSKNLSKILQPVTLGGTSLKHGVTLOT--QAFTOQADSRLEMDVGTLE 482  
 DB 370 THQDDPDQDYGIL--IGSIDQYQNAELNVGSTOOTFVHALTFPONGTLINDAG---- 422  
 QY 483 PADSTINNLVINIS--IDGAKKAKIETK-----ATSKNLTLSGTTLLDPTGTFYE 533  
 DB 423 -----GNVTVNGSFAGIIEAGQLTIAONGSYVLAGAQMALTDIYVDDCAVLSLE 475  
 QY 534 NNSL-----RNOSY-----DILELK--AGTYTSTRAVTPDPI 564  
 DB 476 GDADLITALDDQSDQSYLVNGVLDLSDFTWQSGTSTYNDGEVSGSSGTVIGSDVDVLA 535  
 QY 565 MGEKFHYOGTGWPIY 581  
 DB 536 GGDNLHIGGDKGQYVY 552  
 RESULT 8  
 ID OMPB\_RICJA STANDARD; PRT; 1656 AA.  
 AC 00653:  
 DT 30-MAY-2000 (rel. 39, Created)  
 DT 30-MAY-2000 (rel. 39, Last sequence update)  
 DT 01-OCT-2000 (rel. 40, Last annotation update)  
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)

DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
 DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
 GN OMPB.  
 OS Rickettsia japonica.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=35790;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VH;  
 RA Uchiyama T.;  
 RT "Sequencing of the gene encoding the protein romp B of Rickettsia  
 RT japonica."  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
 CC SIMILARITY).  
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB003681; BAA20138.1; -  
 KM Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1 1338  
 FT DOMAIN 1339 1656  
 FT CONFLICT 528 533  
 FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9D5999F CRC64;  
 Query Match 4.8%; Score 229.5; DB 1; Length 1656;  
 Best Local Similarity 20.1%; Pred. No. 4.3e-05;  
 Matches 188; Conservative 120; Mismatches 321; Indels 307; Gaps 48;  
 QY 22 ATETVLDSSAFQCNKNGNSVRESQEDAGTYLFGKNVLE--NIPTGTAITKSCNNT 80  
 DB 386 ASIAVAITQNSNFTGDBGNLAQVTPD--TWLTGNETGDANNP-----NT 431  
 QY 81 KGDLETFGNNSLTFQTVDA-----VDKSTFTIG-----FSSLSFIAPSGSSIT 100  
 DB 432 AGVITFAANG--TLASASADANAVANTNNTAIEASGVGVQLSGHTRFALRGNAGSVFKL 490  
 QY 101 -GTVAGAAVNSV-----VDKSTFTIG-----FSSLSFIAPSGSSIT 136  
 DB 491 ADGTVINGKNQVIVLGVLAAGAITLDGSAITITGDINGGGAALQSTIYLANDATFTLT 550  
 QY 137 TKGGA--VSCSTGSLKFDKN--VSLFSKN-----FSTDNGAI-----TAKTL 176  
 DB 551 LG--GANLISANGGITINQANGTITKLTSTQNNIVVDCDLAIAIDQTCGVVASSLSTNAQTL 609  
 QY 177 SLGTGWSALFSENT-----SK--KGAIQTSDALITGNGEVSFSDNT----- 220  
 DB 610 TIGG--TIGITGANTTITGQPNISGSKTTLGNGVAINNE--LVIGNNSVGFAPNTIYITR 666  
 QY 221 ---SDSGAALFTEASVTISNNAKVSFIDNKVTGASSSTGDSGCAICAYKTSTDTKVT 277  
 DB 667 TTNAAQGGKIIF--NPVNNNTTLAAGTN--LGSANPLAEINFGS-----KGARADTVL 717  
 QY 278 LTNOMLLESNNTSTTAGAIYVKLELAGGTLPSRNSVNGSTAPKGAIAIEDSGEL 337  
 DB 718 NVEGVNLVAFNTTTTNDAN--VGSFEVFNAGGRNIYS--GTVGQOQGNKFVVAL----- 767  
 QY 338 SLADSQDIY-FLGN--TVSTTPTGNSSIDLGTSAKMTALRSAGAIAIFYDPIITGSS 395

```

CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSDTS AG3/AG5; FUNGI SCT/SC14 AND PLANTS PR-1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 249353; CAA89370.1; -
DR EMBL; X83502; CAA58492.1; -
DR EMBL; X88851; CAA61314.1; -
DR HSSP; P04284; 1CFE.
DR SGD; S0003614; YIT078C.
DR InterPro; IPR001283; -.
DR Pfam; PF00188; SCP; 1.
DR PRINTS; PR00837; V5FPLIKE.
DR PROSITE; PS01009; SCP_AGS_Prl_Sc7_1; 1.
DR PROSITE; PS01010; SCP_AGS_Prl_Sc7_2; 1.
KW Hypothetical protein
PT DOMAIN
SQ
      166      379      ALA/SER/THR-RICH.
      SEQUENCE 881 AA; 89152 MW; 37DBAC660CA9D12A CRC64;

Query Match          4.9%; Score 234.5; DB 1; Length 881;
Best Local Similarity 23.6%; Pred. No. 9,4e-06;
Matches 127; Conservative 84; Mismatches 211; Indels 117; Gaps 22;

QY 39 GNEF--VRESOEDACTTYLEKGN-----VTLENIPG-----TGTATRK 74
DB 112 GHRTQVVMKSTAEICGCGYKCGTMMNNYIVCSYNPNGNLGEFAEEVEPLISVSSSSS 171
QY 75 SCNNNKGDLTFNGNNSLLFQYVDAG---TVAGAAVNSVYDKSTTFIGFSSLSFASP 131
DB 172 SSSSTSTSTSTVST--ISSIMPAVAGYTTTVSSAASSSL--KSTTIPAKATATLASS 227
QY 132 GSSITTKGKAVSCGTSGLFEDKNVSLFKNFSTONGAATAKTLTLGTGMSALFSEMT 191
DB 228 STYTSTSTSGSVSSIVSSASSSV-----TTSYATSSSTYVSSATSTTTTSSVATSSST 283
QY 192 SSKKGAIGTSDALTITNGOGEVFSFDMNTSDSGAIFTEASYTISNNKAVSFIDNKVTG 251
DB 284 TS-----SDP-----TSSPAAASSSDPAAASSAASSSTENAAASSAASSSSSMVSA 332
QY 252 ASSSTTGDMSGAICAYKSTDTKVTLTGNQM--LFSNNTSTTGAAGAIYVKLELASGG 309
DB 333 PLST-----LTTSTASSRSVTSNSVSKFANT----- 361
QY 310 LTFEFSNVGAPAPKGAIAIED--SGELSLSDSGDIYFLAGTNTSTTPGTNRS--SID 366
DB 362 -TVFSAGOTTSVSASASSSSVADDDIGSTSKATSS-----VSHTSTIVSAITAAQYATR 416
QY 367 LGTSAMT--ALRSAGRAIFYDPITGSSSTTVTVLKVNETPADSAIQYGNIIFTGE 424
DB 417 LGSSSSSSGAVSSAVSOSVLSVLAVMVDVSVSSVTAHTTKDTA--TTSVAVASES 473
QY 425 KUSEPAAOS--KNTSLKLPVLTLSGGTSLK--HGVLQTO--ATQOQADSLNDY-- 477
DB 474 ITSETQAOSSTKEKNTSNAATSTSYISNSAVSGHGVYAAAYATTSRQSLATSVPA 533
QY 478 -----GTTLEPADTSTINNLVINISSIDGAKRAKIEFKATSKMLTSGTTLTDP 528
DB 534 TNCSSIVKTKTTLENSSTTTTAT-----KSTTLTATANNSTRAATVATIDPT 582

```

```

DT 01-OCT-2000 (rel. 40, Last annotation update)
DE HYPOTHETICAL 131.2 KDA PROTEIN IN UBIG-NRDA INTERGENIC REGION
DE PRECURSOR.
GN YFAL.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono K., Itoh S., Itoh T., Kanai K., Kasai H.,
RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
RX MEDLINE=84272624; PubMed=6087316;
RA Carlson J., Fuchs J.A., Messing J.;
RT "Primary structure of the Escherichia coli ribonucleoside diphosphate
RT reductase operon."
RL Proc. Natl. Acad. Sci. U.S.A. 81:4294-4297(1984).
RN [4]
RP SEQUENCE OF 925-1198 FROM N.A.
RC STRAIN-K12 / EMG2;
RA Estep P., O'Keefe T., Robison K., Church G.M.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE OF 1180-1250 FROM N.A.
RC STRAIN-OV6;
RX MEDLINE=88201664; PubMed=2834621;
RA Hussain K., Elliott E.J., Salmond G.P.C.;
RT "The parent-mutant of Escherichia coli also carries a gyrAam mutation.
RT The complete sequence of gyrA."
RL Mol. Microbiol. 1:259-273(1987).
RN [6]
RP IDENTIFICATION.
RX MEDLINE=96032851; PubMed=7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes."
RL Nucleic Acids Res. 23:3554-3562(1995).
CC -1- SIMILARITY: TO E.COLI YDEK.
CC -1- CAUTION: REF.3-SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
CC MANY FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000313; AAC75293.1; -
DR EMBL; D90855; BAA16052.1; ALT. INIT.
DR EMBL; D90854; BAA16050.1; ALT. INIT.
DR EMBL; K02672; -, NOT_ANNOTATED_CDs.
DR EMBL; U30459; AAA4094.1; -

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WILMINGTON;
RX MEDLINE=94040787; PubMed=8224886;
RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT "Cloning and sequence analysis of the gene encoding the crystalline
RL surface layer protein of Rickettsia typhi";
RT Gene 133:129-133(1993).
RN [2]
RP PARTIAL SEQUENCE.
RC STRAIN-WILMINGTON;
RX MEDLINE=92114896; PubMed=1370573;
RA Chung W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBR fragments of
RL the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii";
RL Mol. Immunol. 29:95-105(1992).
RN [3]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cleplak W., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RL membrane protein of rickettsiae: Identification of an avirulent
RL mutant deficient in processing";
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; L04661; AAB48987.1;
KM Antigen; S-layer; Transmembrane; Cell wall.
FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
FT TRANSEM 1354 1645 32 KDA BETA PEPTIDE.
FT TRANSEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
FT CONFLICT 657 657 H -> N (IN REF. 2).
FT CONFLICT 842 842 V -> I (IN REF. 2).
FT CONFLICT 1071 1071 G -> A (IN REF. 2).
FT CONFLICT 1306 1306 G -> S (IN REF. 2).
SQ SEQUENCE 1645 AA; 169638 MW; 0CB5641C7EB185EE CRC64;

```

Query Match 5.08; Score 238; DB 1; Length 1645;  
 Best Local Similarity 21.3%; Pred. No. 1.4e-05;  
 Matches 204; Conservative 127; Mismatches 359; Indels 266; Gaps 48;

```

DB 883 TPLDYLTIKSTVGNGVDNFNAPILVSGIDSMIN--NGOVIQDQKN--ITALISGDSN 938
QY 276 VLLTGOMLLEFSNNSTVG-----GATVKKLEIASGL--TEFSRNSVNG 321
DB 939 IIVNSNTLYAGIRTKTGVTLSGGIPNNPGTYIGLENGDRKLKQVTTTDDNNIG 998
QY 322 -----TAPKGAIAIEDSGELSLADSG--DIVELGNTVSTTPGTNRSSID 366
DB 999 SIATNTVITNDVITLTGGLACTDFDKITLGSIGNANVKKVYDFRFSHP-----SMI 1052
QY 367 IGTSAKMTALRSAGRAIYFDPIITGSSVTVDLVKNETPADSALOYTGINIITGEKL 426
DB 1053 VSTKANQ-----GTVTYGNALVGNISSDIPVAVSFNGDSGVGLOGNI-----H 1099
QY 427 SETEADSKNLTSLKLPYTLSSGTLKKGVTLOTQATQOADSRLNDVCTITLEPADT 486
DB 1100 SONIDEGYNLTI--LNSVDILGGGTATNGEIDLTL-----NNLIFANCTSTMGNT 1150
QY 487 S---TINNVIWISSIDGAKKAKIETKAT-----SKNLTLSGTYTL----- 524
DB 1151 SLSTTLNNSGNVGOIVIAEGQVNAATTGTTTITKIDNANANESGTOYTLIOGGARFN 1210
QY 525 -----IDPTGT--EYENHSLRNP-QSY-----DILELKASGTVSTAVYPPDIMER 568
DB 1211 GTLGAPNPDVGTNNIFVKYELIRDANODVYLRTNDVLNV--VTAVGSAIINAPGVHON 1269
QY 569 FHYGQGT-----WGPVWGSGASTATFTFWTKITGIPNERIGS 608
DB 1270 IALCESDVTAAVNNMLAKKSDVATFGAATDTGA--VATVNLNDT-----OKTOD 1322
QY 609 IVPNSLIMNAFIDISSLHLV--METANEG-----LOGDRAP-----WCAGLSNPFHK 652
DB 1323 ILGNL-----GALRTLSNETADVGSSENGVNSGDEALIDQSYGMWAPFNINAEQ 1375
QY 653 DSTKTRGFRHLSSGYVIGNLIHT--CSDKIISAPCQLFGRDRDYVAKNQGTYVGGTLY 711
DB 1376 DKKGGLAGYKAKTAGVAVVG--LDTLANDMLMIGAIGITKTD-----LKHOD-----Y 1421
QY 712 YOHNETYISLPCKLRPCIS--VYPTIEIPVLFSGNLSYTHNDLTKTKTTY----- 761
DB 1422 KKGDT-----DIGLSLSLGAQOLVKNFPAGSAITFLTKVKSRSORFFDANGKN 1475
QY 762 -PTVSGWGNDSFALEFGRAPICDLSALFEQVM--PMKLOFYVAHOGFEKQGT 815
DB 1476 KOIAGNIDN-----ITFGNLMFGYDYNALQGVLTVPAGLSYLSASNENYKETGT 1527

```

RESULT 6  
 XJH8\_YEAST STANDARD; PRT: 881 AA.  
 ID XJH8\_YEAST  
 AC P47033;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HYPOTHETICAL 89.2 KDA PROTEIN IN SCPI60-SM3 INTERGENIC REGION.  
 GN XJ078C OR J1027.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE=96093911; PubMed=7483841;  
 RA Miosga T., Schaaff-Gerstenschlaeger I., Chalwatzis N., Baur A.,  
 RA Boles E., Fournier C., Schmitt S., Veltan C., Wilhelm N.,  
 RA Zimmermann F.K.;  
 RT "Sequence analysis of a 33.1 kb fragment from the left arm of  
 RT Saccharomyces cerevisiae chromosome X, including putative proteins  
 RT with leucine zippers, a fungal Zn(II)-Cys6 binuclear cluster domain  
 RT and a putative alpha 2-SCB-alpha 2 binding site";  
 RL Yeast 11:681-689(1995).

OY 685 ARCQLEGRDRDYFAKNOGYVGGTLYYQHNTEYISLPCKLRPCSLSYPTLEIPVLSGN 744  
 DB 1108 GLLQYOG-----GVANTININLTDN-----ASAVTFPNPVYVTGA 1142  
 OY 745 LSYHTDNDLTKRYTTPYTKGSMGND-----FALEFGG 779  
 DB 1143 INTGNANGNIVTFGTGNTVTDIGTINLATVYVAGITLQAGG 1187

## RESULT 4

SLAP\_CAUCR

ID SLAP\_CAUCR STANDARD: PRT: 1025 AA.

AC P35828; 046015;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE 5-LAYER PROTEIN (PARACRYSTALLINE SURFACE LAYER PROTEIN).

GN R5AA.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

OX NCBI\_Taxid=76;

RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.

RC STRAIN-ATCC 19089 / CB15;

RX MEDLINE=93007489; PubMed=1393820;

RA Gilchrist A., Fisher J.A., Smit J.K.;

RT "Nucleotide sequence analysis of the gene encoding the Caulobacter

RL crescentus paracrystalline surface layer protein.";

RN Can. J. Microbiol. 38:193-202(1992).

RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.

RC STRAIN-ATCC 19089 / CB15;

RX MEDLINE=89080809; PubMed=3049545;

RA Fisher J.A., Smit J.K., Agabian N.;

RT "Transcriptional analysis of the major surface array gene of

RL Caulobacter crescentus.";

RN J. Bacteriol. 170:4706-4713(1988).

RP [3]

RC CHARACTERIZATION.

RX STRAIN-CB15;

RA MEDLINE=98292737; PubMed=9620954;

RT "The Caulobacter crescentus paracrystalline S-layer protein is

RL secreted by an ABC transporter (type I) secretion apparatus.";

CC J. Bacteriol. 180:3062-3069(1998)

CC -I- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY

CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.

CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A

CC S-LAYER WITH HEXAGONAL SYMMETRY.

CC -I- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE

CC SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A

CC SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: AF062345; AAC38665.1; -

CC HSSP: P01549; 2MCM.

CC InterPro: IPR001343; -

CC Pfam: PF00353; hemolysinCabin; 1.

CC PRINTS: PR00313; CABINDNGRPT.

CC Cell wall; S-layer; Calcium-binding.

FT INIT MET

FT 0

FT 0

FT 0

FT 0

FT 0

FT 0

FT 0

FT 0

FT 0

FT 0

FT 0

FT 0

FT 0

Query Match 5.0%; Score 240.5; DB 1; Length 1025;  
 Best Local Similarity 24.8%; Pred. No. 5.4e-06;  
 Matches 183; Conservative 86; Mismatches 253; Indels 217; Gaps 35;

OY 12 TPAIFPLMIATETVIDSSASFDGKNKMFVRESOEDAGTYLFEKGNVLENIPI---G 67  
 DB 268 TVAGEVAGAAATLVGDTLSGAGT-----DVLNMQAAAVPALPGV-VTISGIETMANTV 321  
 OY 68 TGTATTKSCFNNTKGDTLTFTGNKNSLLFQTVDA-----TVAGAAVSVVVDKSTTETI 120  
 DB 322 SGAAITLNTSSGVYTG-LTFLNMTSGAQTITAGAGCONLTATTAQAANNVAVD----- 374  
 OY 121 GFSLSFIASPSSTITGGAVSCSTGSLKFDKNVLSLFESKNESTDNGAITPAKTLSTLG 180  
 DB 375 GRANTYVASTGVTSGTITVGANSAAGTV-----SVSV-----ANSTTTTGAI-----AVTG 422  
 OY 181 TTMASLFESNTSSKKGAIQTSDALTTTGNOC--EVSFSDNTSSDGAALF--TEASVIT 236  
 DB 423 GT--AVTVAQTAGNAVNTLTQTADYVTGNSSPTAVTVQTAAATAGAVRGNCAVTI 480  
 OY 237 SNNA-----KVSFIDKKVYGA-----SSPTGMSG-----GAICAKTSTDT 274  
 DB 481 TDSAAASATTAGKATATVYTGSGAATIDSSALTTVNLSTGTSLGIGRAITA----- 533  
 OY 275 KYTLGNOMILFSSNNTSTTAGAIVYKLELASGGLTFSRNSVNGTAPKG-GAIAIED 333  
 DB 534 --TPRANTLTLMVNGLTTL--GAITDSEAAADG-----FTTINIASSTIASLVAAD 585  
 OY 334 SGEISLSADSGDIV-----FLGNTVYST-----TPGNNRSSIDIGTSA 371  
 DB 586 ATTLNTSGARVYITSHTAALIGIVTNSVGATLGAELATGLVFTGAGRDSILGAVT 645  
 OY 372 KMTALRSAGRAIYFYDPTTGSSTV-----TDVLKVN----- 405  
 DB 646 K-AIYMGAG-----DDIVYVSATLGAAGSVNGDGDIVLAVYVNGSSFSADPAFGF 697  
 OY 406 ET--PADSALOYTGNI--FTGERLSETEA----- 432  
 DB 698 ETLRVGAAGAOSHNANGFTALQATAGATFTVNAVVGTLVLAAPTGTVTLANAT 757  
 OY 433 ---DSKNLTSLKLPVTLGSGTSL-----KHGVLQTOA-----FT 466  
 DB 758 GTSDFENLT--LSSAALAGVVALAGVEVNIATDTMTTAAVDLTLOASAKSIYVT 815  
 OY 467 QOADSRLMDVGTTLLEPATSTI-----NNLVINISSIGAKAKKETATKAT 514  
 DB 816 GAGNGLMTGTGNTAVTSFPAASAVTGAAPVTSANTYGEVYTINGAGADSLTSATA 875  
 OY 515 NLTLL--SGTTLTLDPTGTFYENHSLRNPQSYDILEKASGTVSTAIVPDPPIGKFEHY 571  
 DB 876 NDTIIGAGADPLVYVGT-----DFTGTGADIDFINAIGTSTAVTITDAVAGDKLDL 931  
 OY 572 GYGOTWGPVWGT-GASTT 589  
 DB 932 VGISTNGAIDAGAFGAAYT 950

RESULT 5  
 OMPB\_RICTY STANDARD: PRT: 1645 AA.  
 AC P96989;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROM B)  
 DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
 GN SLP.  
 OS Rickettsia typhi.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiiae; Rickettsia.  
 OX NCBI\_Taxid=785;

```

Db 581 IFNGRLIVDENREYSYDMSDLSRGAEOILSIETTNDDGOLDMSOSLNTSILSPHY 640
QY 572 GYOGTMGPW----- 582
Db 641 GYOLMWP-NMITYTYITLNNSSAPTSAIEOKKTSTFTPSNTTASIPNIKASA 699
QY 583 --GTCAGT-----TATFNTKTGYIPNPERIGSLVENS-----WNAFIDISLHY 626
Db 700 GSGSGSASNSGEVITKHTLVNAPVGIYDPIRRGDLINSLVHSGRNMGLRSL-- 757
QY 627 LMETANGLGDRAFWCAGLSNFFHKDSTKRGFRHLSGGYV-----GGLHTCSDKI 681
Db 758 LPDMSWPAOC-----AATLTFKQOKRLSYHGYSASKGYTSOASGAHG----- 805
QY 682 LSAFCOLFGDRDYPFAKN-----OGTYVGGTLYQHNEITYILPCKLRCSISYPT 736
Db 806 --KFLLSFGSSDCKMEKETNNLSRYYLSALCFE-----PMF 843
QY 737 IPVLFSGNLSTHTDNDLTKYTYTPVKGSGWNSDFALFEGRAPIC-LDESALFEQYM 795
Db 844 DRILILGAACNYGTHMRSFYGKRSKGFHSTL-----GASLNCGLDMSPLRSM 898
QY 796 --PEMKLOFYVAHQEGEQTAREFGSSRL-VNALPFIGIRPKESDC-QDATYVNLTL 851
Db 899 LTPPAQLFSTREPASIRESGDLARLFLBQAHAVVSPIGIKGAYSDDTPTLSWEMEL 958
QY 852 GYVVDLYRSPDCTTTLRISGDSKRTGCTNARQALVLRAGNHCFPNSFEAFESQFSEL 911
Db 959 AYOPTLWKRRRLTLTIQNGSWVTNTPLAKHSFYGR-GSHSLKPSHLTLFANYQAEV 1017
QY 912 RGSRRNYNDGAKYOF 928
Db 1018 ATSTVSHYINAGALVF 1034

RESULT 3
190K RICRI STANDARD; PRT; 2249 AA.
AC P15921;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=783;
RN SEQUENCE FROM N.A.
RC STRAIN=R.
RA MEDLINE=90354033; PubMed=2117568;
RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
RT "A protective protein antigen of Rickettsia rickettsii has tandemly
RT repeated, near-identical sequences.";
RL Infect. Immun. 58:2760-2769(1990).
CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY.
CC -1- PPM: GLYCOSTATED (POSSIBLE).
CC -1- DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF
CC AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: M31227; AAA26380.1; -.
DR PIR: A41477; A41477.
KW Antigen; Repeat; Signal; S-layer; Glycoprotein.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 2249 190 KDA ANTIGEN.

```

```

FT DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 212 286 A (TYPE I).
FT REPEAT 287 358 B (TYPE II).
FT REPEAT 359 430 C (TYPE II).
FT REPEAT 431 505 D (TYPE I).
FT REPEAT 506 577 E (TYPE II).
FT REPEAT 578 652 F (TYPE I).
FT REPEAT 653 724 G (TYPE II).
FT REPEAT 725 799 H (TYPE I).
FT REPEAT 800 874 I (TYPE I).
FT REPEAT 875 949 J (TYPE I).
FT REPEAT 950 1021 K (TYPE II).
FT REPEAT 1022 1093 L (TYPE II).
FT REPEAT 1094 1165 M (TYPE II).
FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
SQ SEQUENCE 2249 AA; 224333 MM; A9D6446C089DF087 CRC64;

Query Match 5.5%; Score 261; DB 1; Length 2249;
Best Local Similarity 22.8%; Pred. No. 1.2e-06;
Matches 202; Conservative 100; Mismatches 325; Indels 258; Gaps 43;

QY 60 VTLENIPGTATITKSCFPNTKGDITFTGN-----GNSLLEFQTVDAQYVAGAAVNSV 112
Db 396 VTFTN-PPVVTGALDNTGNANNGIYTFGTGNTVGDIGNTNALTATVNVG--AGTATILGA 452
QY 113 VDKSTT-----FIGSSLSFTASPPSSITTG-----KGAVSCSGSLKFKKNV 155
Db 453 VIKATTKRLTNAASVLTFTNAAVLTGALDNTGTGDVNLNGALSQVYGNIGNTSL 512
QY 156 SLI-FSKNSTDNGCAITAKTSLTGTMASLFS-----ENTSSKKGAIOTSDAL 205
Db 513 ATISVAGATILGCAVIAKATTKLTDAASAKFTNPVVTGALDNTGNANNGIYTFGNS 572
QY 206 TITGNGE-----VSFSDNTSSDGAII-----FTEA-----SVTISNNAKVSFDN-- 247
Db 573 TVTGDIGNTNSLATISVAGATATLGCAVIAKATTKLTFTNAAVLTFTNAAVLTGALDNTT 632
QY 248 -----KVTGASSSTTGDMS-----GGAICAKYTSPTKTYTLGNQ 282
Db 633 GGDVGVNLNGALSQVYGDIGNTNSLATISVAGATATLGCAVI--KATTKLT-NAVS 688
QY 283 MLFSPNSTTGAAGAIYVKKLELASGGLTFSRNS-----VNGCTAPKG 326
Db 689 AVKFTNPVVT--GAI--DSTGNANNGIYTFGTGNSYTGDIQNTNALTATVNVGAGTATLG 744
QY 327 GAIAIEDSGELS-----LSADSGDIFELGNTVSTTPTGNSSIT-----DLGTS 370
Db 745 GAVIKATTKLTNAAVLTFTNAAVLTG--AIDNTGTGDVNLNGALSQVYGDIGNT 803
QY 371 AKMTALSAAGRAIFYDPITTGSSYTVTVLKYNETPPADALQYT-GNIFTEGKLSSET 429
Db 804 NSLATISVAGATA-----PLGGAIVIAKATTKLT--TNAASVLTFTNAAVLTG-AVDNT 853
QY 430 EADSKNLTSLKLPVTLISGGLTLKHLGVTLOTQAFLOQADSRLEMDVGTLEPADSTI 489
Db 854 TGGDNV-----GVNLN-NCALSQVYGDIGNTNSLATISG-----AGTATL 893
QY 490 NNLVINISSIDGAKKAKIETKATSKNLTLSGTI--TLDPGTGFYENHSIR-----N 539
Db 894 GGAIVIAKATTKLTFTNAAVLT--LTNAAVLTGALDNTGTGDVNLNGALSQVYTGDI 952
QY 540 POSDILEKASGVY--STAVTPRPMGEKFFHYOGTWPVWGCTA----- 586
Db 953 TNSLATISVGA-GTATLGCAVIAKATTKLTDAASAVF-----TNPVV-VTGAIDNTG 1003
QY 587 -----STATFNTKRTGYIPPERI-----GSLVPSINWAFIDISILHYM 628
Db 1004 NANNGIYTFGTGNSYTVGAVGTNALTATVNVGAGLLOVGGVYKANTINTLTNAAVFTN 1063
QY 629 ETANEGLOGDRAFCAGLSN--FFHKDSTKTRGRFRHLSGGYVYVGNLHTCSDKILISA 684
Db 1064 PVVVTG-----AIDNTGNANNGIYTFGTGNSYTV--GNV--GNTNALTATVNVGA 1107

```



OY	223	DGSAIFPEASVITSNNAKYSFIDN-----VTGASSSTDDMGGAICAKTSTDVKVL	278
Db	241	-AGALITSPICSLTGNGCNITFTNNRCPKNETRASSA---DGA- ----KYTRLDV	291
OY	279	TGNO-MLFSNNNTTAGAGAIYYKKLELASGGILTFESRNNGGTARKGALAIEDSGEL	337
Db	292	TGNRGRIFFSDINIRKYNKGAIYAFAVTVLVNDGPYFINNIANN----KGAATYIDGHSN	347
OY	338	SLSADSGDIYLGLTYVSTN--PETH-----RSSIDLCTSAKMPLRSAAARAIFYD	368
Db	348	KISDRHAIIENENIVTNVNANSTSYSANPPRNAIATVAASSGEILLGAGSSONILEFD	407
OY	369	PITGSSSTTVVDYLKAVNETPADSALOQTGNIIFTGEXLISTEADSCKNLTSKILOPYTS	448
Db	408	PIEVSNAGVSYSFNK-----EADQ----TGSVPSGATVNSAD-FHQNKIQTKPAFLTIS	458
OY	449	GCTLSEIKHGVTLOQAATQQAD-----SKLEMDVGTTLEPADDTINNVLVINISS	498
Db	459	NGLFCLIEDHQAOTVNRPTTOGGVVSLNGAVLSOCYKKGSTDSASMA-SILKHIGINTLS	517
OY	499	I--DGAK--KAKEFTKATSKNLT--LSGTLILDPTQTEVENSLRNPOSDYLELKASG	552
Db	518	ILKSGAEIPLLMVEPTNNSNNYADPATATSSLSDVKLSDIDDYGNSPEYESTDLTHALSSQ	577
OY	553	TVTSTAATPD-PIMGEKF-----HYGYOGTWGPIWNG-----TGASTTA- ---591	
Db	578	PMLISEASDNOQSENIDEFSLNVPHYWGQLM---TWGMARTQDEBPASSATIITDPK	634
OY	592	-----FNMTTCGVIIPPERGSLVPRSLM-NAFIDISSLIHYLETANEGLQDRAF	641
Db	635	ANRHRRLULLTWLPAGVVPSPDKRHSPRIANTLWGNMLLATESKNSAELTPSG- --HP	690
OY	642	W---CAGLSNFHDSDKSTRGRGFHLISGCV---VIGNLTSCSDKITLSAFCOLFGRDR	695
Db	691	MGITGGGGMVMVYDDPRENHDFMRBSSGVSAQMINGQHTFEUK-----FSGTYIKME	745
OY	696	YFVAKNGCYTGGILYQHNEITYISLPCKLPCSLSVPLEIFVLES--GNLSYHTDN	752
Db	746	RX-AKNN--VSSKNYCSQGEMLEST-----OEGFLTKILGLVLSYGDHNCHHFYTQ	794
OY	753	DLKRKYTTY--PYTKSGMGMSDFALEFGAPICLDSEALFEQYMFPMKLOFYAHQEGF	810
Db	795	ENLTSQGTFSQTMGAGAVEFDLPKRPGRSHILT-----APFLGALGIYSSLISH	844
OY	811	KEOSTEARERBS-SRLVNLALPGICRFDEKSDCODAT---YNLTGYTVDLVRNSNPC	864
Db	845	TEVCAYPRSESTKPTKLINVLYGVK---GSEFMATTHRQAWTVELAQPVLYROEPGI	900
OY	865	TTTTLRISGDSMKRTGTMALAOALVLRAGN-----HEGCNSNPFAQSOSPFLLRG	913
Db	901	AAQLASKGLWFSGSSSRHAMSYKISQOTPLSWLTLTFQYH-GPYSSTPCNYLNG	958

RESULT 2  
OMPF.CHLTR ID OMPF.CHLTR STANDARD; PRT: 1034 AA.

AC P38008: 084878;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE OUTER MEMBRANE PROTEIN F PRECURSOR.  
GN PMPF OR CT870.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxId=813;

[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=D/UM-3/CX;  
RX MEDLINE=99000809; Pubmed=9784136;  
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R.O., Aravind L.  
RA Mitchell W.P., Olinger L., Tatusev R.L., Zhao Q., Koornik E.V.,  
RA Davis R.W.,  
RT Genome Sequence of an obligate intracellular pathogen of humans:

[illegible]

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2001, 03:32:28 ; Search time 48.09 Seconds

(Without alignments)  
661.033 Million cell updates/sec

Title: US-09-446-677b-10

Sequence: 1 MKSSFPKVFVFFAIFPLSM.....FELRGSSRNRYNDLAKYQF 928

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 segs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	540.5	11.3	964	1	OMPE_CHLTR
2	478	10.0	1034	1	OMPE_CHLTR
3	261	5.5	2249	1	190K_RICRI
4	240.5	5.0	1025	1	SLAP_CAVCR
5	238	5.0	1645	1	OMPB_RICTY
6	234.5	4.9	881	1	YJHB_YEAST
7	231	4.8	1250	1	YFAL_ECOLI
8	229.5	4.8	1656	1	OMPB_RICJA
9	226.5	4.7	995	1	YIO9_YEAST
10	223	4.7	1643	1	OMPB_RICPR
11	222.5	4.7	1609	1	FIG2_YEAST
12	216.5	4.5	2003	1	YDBA_ECOLI
13	216	4.5	1041	1	EGT2_YEAST
14	215.5	4.5	1140	1	YMG6_YEAST
15	215	4.5	1848	1	CBPA_CLOCL
16	214.5	4.5	725	1	AGAI_YEAST
17	212.5	4.4	1300	1	120K_RICRI
18	212.5	4.4	1654	1	OMPB_RICRI
19	209.5	4.4	1120	1	STRF_ECOLI
20	209	4.4	1039	1	AGS4_ECOLI
21	208	4.4	1306	1	MSB2_YEAST
22	208	4.4	1419	1	ALAI_CANAL
23	207	4.3	1569	1	YPIA_ECOLI
24	206	4.3	1286	1	AIDA_ECOLI
25	199.5	4.2	1169	1	YK82_YEAST
26	199	4.2	1150	1	APM0_PIG
27	197.5	4.1	733	1	SLAP_CAME
28	193.5	4.0	749	1	TROP_HUMAN
29	193.5	4.0	796	1	YSBA_CABEL
30	190.5	4.0	959	1	N100_YEAST
31	189	4.0	1251	1	Y003_CABEL
32	188	3.9	1322	1	ICBA_PANAN
33	185.5	3.9	610	1	MUCA_HUMAN

34	185.5	3.9	774	1	STF_LAMB	P03764 bacterioph
35	185.5	3.9	1577	1	HUYA_PROMI	P16466 proteus mir
36	185	3.9	598	1	YAU6_SCHPO	010168 schizosacch
37	181.5	3.8	3591	1	FHAB_BORPE	P12255 bordetella
38	180.5	3.8	2090	1	N214_HUMAN	P35658 homo sapien
39	179.5	3.8	1325	1	YDEK_ECOLI	P32051 escherichia
40	177	3.7	797	1	VGLX_HSEVB	P28968 equine herp
41	176.5	3.7	1367	1	AMYL_YEAST	P08640 saccharomyc
42	175	3.7	1161	1	YJ9P_YEAST	P47179 saccharomyc
43	174	3.6	1258	1	ICEN_ERMHE	P16239 etwilia her
44	173.5	3.6	1567	1	ICEN_XANCT	P18127 xanthomonas
45	173	3.6	863	1	YEJO_ECOLI	P33924 escherichia

## ALIGNMENTS

RESULT	1	STANDARD	PRT	964 AA.
OMPE_CHLTR	OMPE_CHLTR			
ID	OMPE_CHLTR			
AC	084877			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	PUTATIVE OUTER MEMBRANE PROTEIN E PRECURSOR.			
DE	PMPE OR CT869.			
GN	Chlamydia trachomatis.			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI-TaxID-813;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-D/UM-3/CX;			
RX	MEDLINE-9900809; Pubmed-9784136;			
RA	Stephens R.S., Kalman S., Lammell C.J., Fan J., Marathe R., Aravind L.,			
RA	Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,			
RA	Davis R.W.;			
RT	"Genome sequence of an obligate intracellular pathogen of humans:			
RT	Chlamydia trachomatis."			
RL	Science 282:754-759(1998).			
CC	-I- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)			
CC	(POTENTIAL).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL: AF001360; AAC68467.1; ..			
KW	Outer membrane; Signal.			
FT	SIGNAL			
FT	CHAIN			
FT	SEQUENCE			
FT	62			
FT	68			
FT	118			
FT	123			
FT	169			
FT	181			

Query Match 11.3%; Score 540.5; DB 1; Length 964;  
Best Local Similarity 25.4%; Pred. No. 1.4e-22;  
Matches 244; Conservative 138; Mismatches 402; Indels 175; Gaps 46;

QY	62	LENIPTGTAT-----KSCFNTKGDLEFTGNSLTFQYVDACVAGAAVSSVVDKST 117
DB	68	LQKTPNCAVATITDIYLSFPTQKGIYFAKN-----LTPSSGAIGYASNSFTVELRD 122
QY	118	TF--IGFSS-----LSFIASP--GSSITGKAVSCSTGLKFDKNVSLFSKNFTDNG 168
DB	123	TIGVIEFNNTCCRLFWMRNRYADKIREG-GALHAQNLVYINHHQV-VGFMKNFSTYQG 180
QY	169	GAT-TAKTSLITGTTGALFSEN-----TSSKKGAIQTSDALTTITNGQGVSSSDWTS 222
DB	181	GAISTANTVFVSENOGFLFMDNICIOTNTAGKGAIVAGTSNFSFESNMCDLFIINNACC 240

QY 762 PIVKSGWGDSDALFEGRAPICLDESALEFQYMPFMKLOFYAHOGEFQGT-ERREF 820 ?  
 DB 678 SEIKGDWGDCEGVEFGAKAPI-ETASILFDMYSPFKQLVNHQDDEENNDDGRF 736  
 QY 821 GSSRIYVIALPFGIRFDESDODATYTLTGTYVDLVRSNPDCTTLRLISGDS--WKTF 878  
 DB 737 EENNLTJNSMPLGYKLEFESHKDTASYNLTLAYAPDIVRSNPDCTASLVSPTSAYWTK 796  
 QY 879 GTNARQALVLRAGNHCFNSFEAFSQFSELRGSSRNRYNDLGAKYOF 928  
 DB 797 ANNLAHRAFLIAGNVLALFRNTELFSGFELRGSCRTYNDLGSRIOP 846  
 RESULT 15  
 Q923A1 PRELIMINARY; PRT; 841 AA.  
 ID Q923A1  
 AC Q923A1  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE OUTER MEMBRANE PROTEIN 7 (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).  
 GN OMP7 OR PMP-2 OR CP0761.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.  
 OX NCBI\_TaxID=83358;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CML029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalan S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Gilmwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VRI310;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Christensen G., Birkelund S.;  
 RT "Chlamydia proteins containing the GCAT-repeat belong to a subfamily  
 of autotransporting pathogenicity factors.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Reed T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouil H., Craven B.,  
 RA Bowman C., Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DB EMBL: AEO01586; AAD18172.1;  
 DB EMBL: AJ133035; CAB37083.1;  
 DB EMBL: AE002235; AAF38561.1;  
 DB TIGR: CP0761;  
 DR InterPro: IPR003357;  
 DR InterPro: IPR003368;  
 DR Pfam: PR02385; OMP\_1;  
 DR Pfam: PR02415; DUF145; 1;  
 SO SEQUENCE 841 AA; 89600 MW; 9064D60D67BD24C CRC64;

Query Match 32.7%; Score 1562; DB 2; Length 841;  
 Best Local Similarity 39.2%; Pred. No. 1.4e-81;  
 Matches 369; Conservative 121; Mismatches 312; Indels 140; Gaps 20;

QY 12 TRAIFPLSLATEVYVDSASPDG-NKNGNFSVRSQEDACTYVLEFGNVTLENIPTGT 70  
 DB 15 TUSMSNLGAATTELSASNSPDGTTSTTSSTSSATDGTNTVFKDSVYIENVPKGE 74

QY 71 AITKSCFNN--TKGDTFTGNGNSLLFQTVDAGYAAGAAVSVVXKSTTFIGSSLSFI 128  
 DB 75 TOSTSCFKNDAAGLNLNLCGGFSFTFENIDATPAGSAIGSEANATVLLSGFALSFL 134  
 QY 129 ASPGSSITTGKAGVSCGSLKFPKNSVLSFSKNFSTDNGAATKATLSLGTMTSLFS 188  
 DB 135 KSPASTVINGALINVK-GNLSLDNDKVLQDNFSTGDDGALNC----- 178  
 QY 189 ENTSKKGGALQTSBALITTONGEVSESDNTSSDGAALFTFASVYISNNKAVSFIDNK 248  
 DB 179 -----AGSLKIANKNLSLFI--- 193  
 QY 249 VTGASSSTTGDMGSAICAVKSTDTKVTLTGNOMLFSNNSTTGAGAIYVKKLELAG 308  
 DB 194 -----GNSSTTGAGAIYVKKLELAG 214  
 QY 309 GLTLFSRNSVNGCTAP-----KGAIAIEDSGELSLSDSGDIVFLGNTVSTTPTG-NRS 363  
 DB 215 GETLF-----QGNAPAPMAKGGAIADSGTSLISGSDGDIIFEGNTIGAT--GVSHS 267  
 QY 364 SIDGTSKAKTALRSAGRAIYFYPDTTGGSTTVTDVKKVNETPADSLQTTGNIITFG 423  
 DB 268 AIDGTSKAKITLRAAQGHITFYFDPITVGTSTVADALNINSPDTGDKKEVGTIVSG 327  
 QY 424 EKLSETEADSKLTLKTLQPTVLSGCTLSLKHGYTLQTAFTQOADSRLXMDVGTTEP 483  
 DB 328 EKLTEBAKDEKRTSKLLQNAVAFKNGYVVLKGDVLSANGFSDQANSKLLMDGTSI-V 386  
 QY 484 ADTST--NNLVINISSIDAKKAKIETKATSKNLTSLITLDPPTFENSLRNPQ 541  
 DB 387 ANESIEIENLEINIDSLNGKIKLSATQAKDRIDRPVVALISDSFQNGFLENEDH 446  
 QY 542 SYD-IETKASGVSTANTPPPIGKEFHGYOGTWGPPIVWGASTATFATNMTKTYI 600  
 DB 447 SYDILLEDAGKDIVISDSRS-IDAVOSPYOGKW-TINWSTD-DKKAIVMAKQSFN 503  
 QY 601 PNERIGSLVPSNLNNAVIDISLHYLMETANEGLOGDRAFCAGLSNFFHKDSTKTRG 660  
 DB 504 PFAEBAFLVPLPLWMSPTIDVRSKONFTELTGEPYKRWVAGISNVILHRSRGRENQK 563  
 QY 661 PRLHSGGVIGGNLHTGCDKTLISAFQCLFGRDRDYPAKNOGTVYGGTYOHNETIYS 720  
 DB 564 PRLHSGGVIGGNLHTGCDKTLISAFQCLFGRDRDYPAKNOGTVYGGTYOHNETIYS 720  
 QY 721 LPC-----KLPCSLSYVPEIPLVPSGNLSTHTNDLTKY-----TTPYVKGSGW 769  
 DB 624 VVSTLLGRGGLREILLPVYKTLPCSYGQSLSTGHDRMKTESLPPPTLSTLSTHSTW 683  
 QY 770 NDSALREGGAPICLDESA--LEQYMPFMKLOFYAHOGEFQGTAREFGSSRLV 826  
 DB 684 GYVAGELGTR--VAVENTSGRGFEYTPFVKVQAVYARQDSFVELGALSRDFSSHLY 741  
 QY 827 NLALPIGRFDESDODATYNTLGTVDLVRSNPDCTTLRLISDSSKKTGTNLARQA 886  
 DB 742 NLALPIGRFDESDODATYNTLGTVDLVRSNPDCTTLRLISDSSKKTGTNLARQA 886  
 QY 887 LVLRAGNHCFNSFEAFSQFSELRGSSRNRYNDLGAKYOF 928  
 DB 800 GIVQASGFSRLGAALIELFNGFEGMGSSRSYVNDAGSNIKF 841

Search completed: October 2, 2001, 03:31:33  
 Job time: 273.sec

```

Db 186 EN-----NONVFAENSSSSSG 203
QY 297 AIYKLELASGGLTFRSNVNGTAPKGAIAEDS-GEISLADSGDIYVLTGTVS 355
Db 204 AIYKLELASGGLTFRSNVNGTAPKGAIAEDS-GEISLADSGDIYVLTGTVS 262
QY 356 T---TPGTNRSTLGTAKMTALRSAAARAYEPDPTTGSSTTVYDLKVNTPADSA 412
Db 263 TNGSGPTVNRNSIDGSSGKFTKMAKEGCIFFYDITGGGS---DELINIK---QDT 315
QY 413 LQYTNIIIFTEGKLESEPADSKNLTSLKLOPTVLSGGLTSLKHGVTLOQAFOQADR 472
Db 316 VDYTGKIVFSGRLSDERKVAANLKSDFOPLKIGSSSLIKGGLVLEKFSFQITGAT 375
QY 473 LEMDVTLEP-----ADSTINNLVINISIDG-----KKAIEFKATSKNLSITTL 524
Db 376 VYMDLGTLLQPPSSGGTITLLNDINVASLGGGVADPAKVEATETSKVTIN-AVNL 434
QY 525 LDPTGTFYENHSLRNPDSYDILEL-ASGVTVSTAV-----TPDPIGKEFHYGYOGTW 577
Db 435 VDDNGNAYVPIILASQPTALEVRSGSGSITKPTTLENYTPPT-----HGYOGNW 488
QY 578 GPTVGTGAST---TATFWTKTGYPINPERIGSLVPSNMAFIDISLHMETANEG 634
Db 489 -TWTNKGSSADERTYATLWEOTGSPNEROGSLVPTNLMGFSFDIRAIONLMDISYNG 547
QY 635 LQGDRAFWCAGLSNPFHDKSTKTRGFRHLGGYVIGNLHTCSDKLSAARCOLFGDR 694
Db 548 ADYHGRFVWSGLNPLHKSQSDTKRFRHNSAGYALGYAQPSEDFVSAACQLGKDK 607
QY 695 DYFAKNGTVYVGYLYOHNFTYISLPCKLRCSLSVPT-----EIPVFSG 743
Db 608 DYFAKNGTVYVGYLYOHNFTYISLPCKLRCSLSVPT-----EIPVFSG 652
QY 744 NLSTHTDNDLTKRY-----TYPTVKGWNGDSFALFEGGRAPICLDEALFEOM 795
Db 653 OLATCHSNKMTMTYAPPKTYSEIRKGDWNGDCEFGVFAKAPIT-ETALFEDMS 711
QY 796 PFMKLOFYVAHQEFKROGT-EAREFGSSRLVNLALPIGIRFESDODATVNLTYGT 854
Db 712 PFMKLOFYVAHQEFKROGT-EAREFGSSRLVNLALPIGIRFESDODATVNLTYGT 771
QY 855 VDLVSNPDCITTLTISGDS--WKTFTNLARQALVIRAGNHCFNSNFAFOESFELR 912
Db 772 PDIVASNDGASLVSATVWTKANILARHAFIIQAGNYTALTRNTELSOFGEELR 831
QY 913 GSSRNYNDLGAQYOF 928
Db 832 GSCRTYNNIDLSKIOF 847

```

```

RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RA MEDLINE-98187897; PubMed-9529048;
RX Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RT "Molecular cloning and characterization of the genes coding for the
RT highly immunogenic cluster of 90-kilodalton envelope proteins from the
RL Chlamydia psittaci subtype that causes abortion in sheep.";
DR Infect. Immun. 66:1317-1324(1998).
DR EMBL: U65943; AAC15923.1; -.
DR InterPro: IPR003357; -.
DR InterPro: IPR003358; -.
DR Pfam: PF02385; OMP; 1.
DR Pfam: PF02415; DUF145; 1.
DR Signal.
KW SIGNAL.
FT CHAIN 1 16 POTENTIAL.
FT SIGNAL 17 846 POMP91B.
SQ SEQUENCE 846 AA; 90834 MW; 4CDDC31D0C3C2964E CRC64;

Query Match 33.5%; Score 1598.5; DB 2; Length 846;
Best Local Similarity 38.6%; Pred. No. 1,1e-83;
Matches 367; Conservative 139; Mismatches 271; Indels 173; Gaps 26;

QY 23 TETVLDSSASFDGN-KNGNFSVRESOEDACTYLFKGNVTELENIPGTATIKSCFNNTK 81
Db 26 TNETLTSDDSYNGNVTSDSEVEKET--TSGATYTCBNVCI-SYAGKDSPLKSCFSEFT 82
QY 82 GDLFTFGNGSLFQRYVDACTVGAAVNSVVDKSTTFIGFS--SLSTASGSSITGK 139
Db 83 ENLSFENGTYLCEPDNITTSQSHPGALSVSGTKNTKLDIGSFLSCAYCPCPG--TGY 139
QY 140 GAVSCSTGSLFKDNVSLFSKNESTDNGATITAKLSLGLTMSALFSENTSSKKGAI 199
Db 140 GAIO-TKGTTLKKNSSLVPHKNCSTAGALOCK----- 173
QY 200 QTSOALITTGQGVSPSDMTSSDGAIFTEASVYLTISNAKVSFDNKVYTAGSSSTGD 259
Db 174 -----SSSTAE 180
QY 260 MSGAICAYKTSTDKYVTLTGOMLTSNNTSTAGAIYVKKLELASGLTFRSNV 319
Db 181 LK-----LENNKMLVSENSSEKKGAIYADKLTIVSGGPTLFSSNSVS 224
QY 320 GGTAPKGAIAIEDS-GEISLADSGDIYVLTGNTVST--TPGTNRSSIDLGTSAKMTA 375
Db 225 HNSSPRGALICKDDGCSILANGLITFDGKILITNGSSPVTNNSIDLGSGGFTK 284
QY 376 LRSAGRAIFYDPI-TGSSSTTVYDLKVNTPADSALOYTGNITFTEGKLESEPADS 434
Db 285 LNAKEGFIFFYDPIANTNGSTPEI---ELNKTESDIT--YTGKIVBSGELSDEKTVP 338
QY 435 KNLTSKLDOPYVLSGGLTSLKHGVTLOQAFOADRLENDVGTLE---PADSTIN 490
Db 339 ANLKSRYKOPKIKGASLVKKGVTLEAKKITQGTGVYMDLGTTLQTPSSGETITLT 398
QY 491 NVVINISSI---DGAKKAKIETKATSKNLTSLGTTLLDPTCTFEYENSLRNPDSYLE 547
Db 399 NLDININLASLGGGGTAPKATATNTASQAISA-AVNLVNTDSNTEYEDILASKFSAI- 456
QY 548 LKASGVTVSTAVTPDPIGKEF---HYGYOGNCPVWGTGAST---TATFWTKTYI 600
Db 457 ---TATSSSTVTPPETNKNKYTPYHIGQGNW-TVTNKGSSAQEKRAVLTWQGTGS 512
QY 601 PNPERIGSLVPSNMAFIDISSLHMETANEGLOGDRAFWCAGLSNFAFHDSKTRG 660
Db 513 PNPERVGSVLPNTLMGAFSDTRAIQNLMDISVNGADYSKGFVWSILANLMSGSDTKR 572
QY 661 FRLHSGGVYIGNLHTCSDKLISARCOLFGRRDVFVAKNGTVYGYLYOHNFTYIS 720
Db 573 FRHNSAGYALVAYQTPSDVCSAFCOLKDKDFVSNNSITIASGYOYH----- 626
QY 721 LPCKLRCSLSVPT-----EIPVLFSGNLSYTHTDNLTKTYI-----TY 761
Db 627 -----ISYNTVNTVTLQNTLGAAPLVNLAQLTVCHASNMMKNTMTNTYTPRNVTP 677

```



Db 462 DSGTLLSTAGSITITNGLNVDISLGLKQPVSLTAKGASNKYVSGKLNLDIEGNIYES 521  
QY 535 HSLNPOSIDLELKASGTYS----TAVTPDPIMGEFHHGYGOTWPIYWGICASTT- 589  
Db 522 HMFSDHDLFSLLKTTVDADVDTNVDISSLIPVPAEDPSEYGFQGOQN-VMMTTDTAINT 580  
QY 590 --ATFNMTKTYINPERIGSLVPSNLNNAFIDISLHLYMETANEGLQGRAPFCAGLS 647  
Db 581 KEARATWTKTFVSPSEKSAVLCNTLMGVFTDIRSLQQLVEIGATGHEHKGFEVSSMT 640  
QY 648 NEFHKDSKTRRGRFRLHSGYVIGNLHTCSDKILISAFCOLFGHRDYPVAKNOGTYYG 707  
Db 641 NFLHKTGDENKGRFRRHTSGYVIGSANTPKDDLTFAFCHLFAADKOCFAHNNSSRTYG 700  
QY 708 GTLYOHNET-----YISL-PCKLPCSLVYPIEIPVLFSGNLSTYTHDLDKTKYTY 761  
Db 701 GTLFFKSHHTLOPONYLMLGRKFSSESAIEKPREIPLALDOVSFSHSDRMETHYSL 760  
QY 762 PTVGSGWGNDSFALFEGGRAPICL-DESALEQYMPFKLOFYVAHOGKFEQGEAREEF 820  
Db 761 PESGSGMNECIAGIGLIDLPVLSNPHLPKTFIPQKKEVYVYSONSFESSSDGRGF 820  
QY 821 GSSRLVNLALPIGIRFDEKSCODA-TYNLTGLYTVDLVRSNPDCOTTLRLISGDSWKTEG 879  
Db 821 SIGRLNLTIPVGAKE-VQGDIGDSYTYDLSGFFVSDYVRNNPOSTATLVMSPSWKIRG 879  
QY 880 TNLARQALVLRAGNHFCPSNFEAFSOFSEFLRGSSRYNVDLGAKEY 928  
Db 880 GNLRSQAFLLRGSNNYVNSNCELFGHYAMELRGSSRYNVDVGTKLRF 928

RESULT 11  
Q9K299 PRELIMINARY; PRT: 949 AA.  
ID Q9K299  
AC Q9K299;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)  
DE POLYORPHIC MEMBRANE PROTEIN G FAMILY.  
GN CF0302.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brubham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,  
Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
Bowman C., Dodson R., Gwin M., Nelson W., Deoxy R., Kolonay J.,  
McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia  
pneumoniae AR39."  
RL Nucleic Acids Res. 28:1397-1406(2000).  
DR EMBL: AE002192; AAF38159.1; -  
DR TIGR: CP0302; -  
DR InterPro: IPR003357; -  
DR InterPro: IPR003368; -  
DR Pfam: PF02385; OMP; 1.  
DR Pfam: PF02415; DUF145; 1.  
SQ SEQUENCE 949 AA; 101357 MW; A00B09E16C699BE3 CRC64;

Query Match 38.8%; Score 1853; DB 2; Length 949;  
Best Local Similarity 42.6%; Pred. No. 3.6e-98;  
Matches 404; Conservative 171; Mismatches 332; Indels 42; Gaps 20;

QY 1 MKSFFPFFVSTRIFP--LSMATEVYLDSSASFQGN-KKGNSSVRESQDAGTYYLFK 57  
Db 22 MKTSIPVYLVSVAFSCHLOSLNDELISPDSDFNNGNIDSGFTPTPTS---ATTYSLT 77  
QY 58 GNVYLENIPGTAITSCFNNTKGDLLFTNGNSLLFQYVDACTVAGAAVSSVVKST 117

Db 78 GDVEFFE-PGKGTPLSDSCFQOTMDNLFPLGNGHSLFRGFIADGTHAGAAA-STTANKNL 135  
QY 118 TPIGSSLSFTASPOSSITTTGKAVSCSTGSLKEPKNVSLEFSKFEYSDNGAITAKLS 177  
Db 136 TFSGSLSPSSPSTYVTOGGLS-SAGVNLNIRKLIVAGNFSTADGALKAGSFL 194  
QY 178 LMGTHMALFSPNTSSKKGALQTSDALITGNOGEVFSNDTSDSGAALFTEASVTS 237  
Db 195 LMGTHMALFSPNTSSKKGALQTSDALITGNOGEVFSNDTSDSGAALFTEASVTS 254  
QY 238 NNAKVSFIDNKVVTGASSTTGDMSGAICAAYKSTDTKVTLLTGOMLFSNNTSTTAGA 297  
Db 255 NKKLFY-----EGMAKTT-----GGAICNTRKASGSPELLISNNKTLTFAVNAETSGA 305  
QY 298 IYKLELASGGLTLPNSNVNGTAPKGAALAEDESELSDSGDIYVLGNTYST- 356  
Db 306 IHAKKIALSSGFEFLFNNSVAT-PKGAISIDASELSLSAETGATFVRNLTLYTG 364  
QY 357 -TPGTRSSIDIGSAKMTALRSAGRAIFYDPITTSSTVTVDLVKNETPADSALOY 415  
Db 365 STDPKRNAINIGSKGFTLELRAKNHTTFYDPT--SEGTSDVLKINNGSAGALNPY 422  
QY 416 TGNITFTGEKLESEADSKMLTSLKLOPVTLSGGTSLKRGVLTQTAFTQOADSRLM 475  
Db 423 OGTLIFSGEITLADBLKADMLKSFQTPVSLSGKLLQGVLTLESFSQOAGSLGM 482  
QY 476 DVGTTLE-PADSTNNLVINISIDGAKKAKIEFKATSKMLTSGTTLTLDPTGFLEN 534  
Db 483 DSGTLLSTAGSITITNGLNVDISLGLKQPVSLTAKGASNKYVSGKLNLDIEGNIYES 542  
QY 535 HSLNPOSIDLELKASGTYS----TAVTPDPIMGEFHHGYGOTWPIYWGICASTT- 589  
Db 543 HMFSDHDLFSLLKTTVDADVDTNVDISSLIPVPAEDPSEYGFQGOQN-VMMTTDTAINT 601  
QY 590 --ATFNMTKTYINPERIGSLVPSNLNNAFIDISLHLYMETANEGLQGRAPFCAGLS 647  
Db 602 KEARATWTKTFVSPSEKSAVLCNTLMGVFTDIRSLQQLVEIGATGHEHKGFEVSSMT 661  
QY 648 NEFHKDSKTRRGRFRLHSGYVIGNLHTCSDKILISAFCOLFGHRDYPVAKNOGTYYG 707  
Db 662 NFLHKTGDENKGRFRRHTSGYVIGSANTPKDDLTFAFCHLFAADKOCFAHNNSSRTYG 721  
QY 708 GTLYOHNET-----YISL-PCKLPCSLVYPIEIPVLFSGNLSTYTHDLDKTKYTY 761  
Db 722 GTLFFKSHHTLOPONYLMLGRKFSSESAIEKPREIPLALDOVSFSHSDRMETHYSL 781  
QY 762 PTVGSGWGNDSFALFEGGRAPICL-DESALEQYMPFKLOFYVAHOGKFEQGEAREEF 820  
Db 782 PESGSGMNECIAGIGLIDLPVLSNPHLPKTFIPQKKEVYVYSONSFESSSDGRGF 841  
QY 821 GSSRLVNLALPIGIRFDEKSCODA-TYNLTGLYTVDLVRSNPDCOTTLRLISGDSWKTEG 879  
Db 842 SIGRLNLTIPVGAKE-VQGDIGDSYTYDLSGFFVSDYVRNNPOSTATLVMSPSWKIRG 900  
QY 880 TNLARQALVLRAGNHFCPSNFEAFSOFSEFLRGSSRYNVDLGAKEY 928  
Db 901 GNLRSQAFLLRGSNNYVNSNCELFGHYAMELRGSSRYNVDVGTKLRF 949

RESULT 12  
P77792 PRELIMINARY; PRT: 839 AA.  
ID P77792  
AC P77792;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)  
DE POWP90B PRECURSOR.  
OS Chlamydia psittaci (Chlamydia psittaci).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83554;  
RN [1]  
RP SEQUENCE FROM N.A.





DT 01-MAR-2001 (TReMBLrel. 16, last annotation update)  
DE OUTER MEMBRANE PROTEIN 11 PRECURSOR.  
GN OMP11 OR PMP 8.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.  
OX NCBI\_taxid=83558;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RL Nat. Genet. 21:385-389(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VR1310;  
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,  
RA Christensen G., Birkelund S.;  
RT "Chlamydia proteins containing the GGAT-repeat belong to a subfamily  
RT of autoantigenic pathogenicity factors.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE001627; AAD18590.1; -  
DR EMBL: AJ133034; CAB37068.1; -  
DR InterPro: IPR003357; -  
DR InterPro: IPR003358; -  
DR Pfam: PF02385; OMP; 1.  
DR Pfam: PF02415; DUF145; 1.  
KW Signal.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 930 OUTER MEMBRANE PROTEIN 11.  
SQ SEQUENCE 930 AA; 97669 MW; 46A9B5E3B913C4C CRC64;

Query Match 40.3%; Score 1925; DB 2; Length 930;

Best Local Similarity 43.3%; Pred. No. 2.6e-102; Matches 413; Conservative 164; Mismatches 328; Indels 48; Gaps 17;

QY 1 MKSFPKVFSTFAIFPLSM-IAT---ETVLDSASFQDKNGKNGFVSRESQEDAGTYL 56  
DB 1 MKTPLKLLLSLTVTYILSLIATYTGADASLSPDSDFGAGSGFTTFRKSTRDANGTYVL 60  
QY 57 KGNVTLNIGTGTAITKSCFNMTKGLTFNGNSLLFQTVADGTYAAGAVNSVYDKS 116  
DB 61 SGNVYI-NDAGKGTALHGCFTETTGDTLFTGKGYSPFTVAGSNAAGAA-STADKA 118  
QY 117 TPTGSGSLSPFASPGSITTGKAVCSGSLKFDKNVLSLKNFTD---NGCAITA 173  
DB 119 LFTGFSNLSFIAAPGTTVAGSKTSL-SAGALNLDNGTILFSSQVNSNEANNNGAITT 177  
QY 174 KTLSTLGTMSALFSENTSSKKGAIOTSDALITITGNOGVSPSDNTSSDGAIFTEAS 233  
DB 178 KTLSTISGNTSITFTYSNAKLGALISSAASISGNTGOLVPMNNKGEMGGALFEAS 237  
QY 234 VTISSNAKVSFIDNKYVAGASSSTTGDMSSGGAICAYKTSTDTKTYLNGOMLFSNNTST 293  
DB 238 SSITQNSLSEFGSNTMTADAG-----KGAICYCEKGTGEPITLISGNKSLFEAENSVT 291  
QY 294 AGAIIYKKIELASGGLTLEFSKNSVNGTAPKKGALAIEDSGELISADSGDLYFLGNTV 353  
DB 292 QGGALCAHAGLDLSAAGPTLPSNNKCGNTAAGKGAALAIADSGISLSANOGDITFLNLT 351  
QY 354 TSTT-EGTNSSIDLGTSAKMTALRSAGRAIYEPDI---TTGSSSTTVYDVLKVNTPA 409  
DB 352 TSTSAFTSTRNALYISSAKITMLRAOGOSITFYDPIASNTGAS---DVLITNPDS 407  
QY 410 DSAIQTGNIITFTGKLESEADSKNLTSKLPVLSGGLTSLKHGYTLQTOAFQQA 469  
DB 408 NSPLDYSGLTIVFSGEKLSDAEKADNFSTILKQPLALASGTLAKGNVELDVNGFTQTE 467  
QY 470 DSHLEMDVGTLEPATDSTIN--NLVINISSIDGAKKAKETKATSKNLTLSGTITLLDP 527  
DB 468 GSTILMQPGTKL-ADTEAISTLTKLVVDLSALEGNKSVSIEGTAGANTITLTSPLVFODS 526

QY 528 TGTVEHNSLNRNPOSYDILELKA-----GTVSTAVTPDPIMGEKPHYGOGTWG 578  
DB 527 SGNFESTITINQATOPLYVTATTAASDIYIDALISPVTPPEP-----HYGQGHME 580  
QY 579 PIWVGASTATENMTKGTGYPNPERIGSLVPSLWNAFIDISSLHYMETANEGLQGD 638  
DB 581 ATWADTSTAKSGTWMTWYTTGYNPNPERRASVAPDLSASFIDITLQOIMTSQANSIYQ 640  
QY 639 RAEFCAGISNFHKDSIKTRGRFHHLSGQYIGGNLHTCSKIISAAPCOLFGDRODYV 698  
DB 641 RGLWASGTANFFHKDSIQNOAFRHKSYGLVGSADDFSENISVAFCOLFGDKDLFI 700  
QY 699 AKNOGYGGTLYYQHNETYISLPCKLRPC--SLSYPTETLPEVFSGLNSTRHDNDUKT 756  
DB 701 VENTSHNTLASLYLQHRNFLGGLP--MPSFGSTIDMKDLPILNLQOLSTYTKNDMDT 757  
QY 757 KYTTPYTVKSGWSDSPALFEGRAPICL-DESALFEOYPMFKLOFYVAHQEGFRKEDGT 815  
DB 758 RYTSYPEAGSMTNNSGALDELGSLALYLPKEAPFGQYPPFLFAQVYSNOQFKESGA 817  
QY 816 EAREFGSSRLVNLALPTGIRPKESDCQDATYNTLTGYTVDLVNSNPDCITTLTISGDSW 875  
DB 818 EARAFAFDGDLVNCISIPYGRLEKISEDEKNNFEISLAVIGDYVKKRNPRTSLWVSGASW 877  
QY 876 KTEGTNLARQALVLRAGNHCFNENFEAFSOFSELGSSRNRYVDLGAQYQF 928  
DB 878 TSLCKNLARQAFASAGSHLTLSPHVELSGEAYELGSAHITVNDGCLRXSF 930

RESULT 9  
P71135 PRELIMINARY; PRT; 926 AA.

DT 01-FEB-1997 (TReMBLrel. 02, Created)  
DT 01-FEB-1997 (TReMBLrel. 02, last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, last annotation update)  
DE PUTATIVE 98 KDA OUTER MEMBRANE PROTEIN.  
OS Chlamydia psittaci (Chlamydia psittaci).  
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.  
OX NCBI\_taxid=83554;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OVINE ABORTION S26/3;  
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U72499; AAB18188.1; -  
DR InterPro: IPR003357; -  
DR InterPro: IPR003368; -  
DR Pfam: PF02385; OMP; 1.  
DR Pfam: PF02415; DUF145; 1.  
SQ SEQUENCE 926 AA; 98439 MW; 3E755E2F594750F CRC64;

Query Match 39.9%; Score 1908; DB 2; Length 926;

Best Local Similarity 42.8%; Pred. No. 2.4e-101; Matches 409; Conservative 155; Mismatches 335; Indels 56; Gaps 18;

QY 1 MKSFPKVFSTFAIFPLSM-----IATETVLDSASFQDKNGKNGFVSRESQED 50  
DB 1 MRSPLYKILSLSTLPLTSPHFSQLHAELVLTQESILDA-----NGAFSPQSTSTAG 52  
QY 51 GTTYLTFKGNVTLNIPGTGTAITKSCFNMTKGLTFNGNSLLFQTVADGTYAAGAVNS 110  
DB 53 GTTYNESDISIVDV-GQTAALASAFVQTDNLTTFKGNHNSLSTNANG-ANPAIINV 110  
QY 111 SYVDKSTTFYFGSLFPIASPGSITTGKAVCSGSLKFDKNVLSLKNFTDNGCA 170  
DB 111 NTADKILITLDFEKLSPKEPSSLYNTKGAMK-SGGALNLANNASTLFPQNTSAENGCA 169  
QY 171 ITAKTSLTGTYSALFSENTSSKKGAIOTSDALITITGNOGVSPSDNTSSDGAIF 230  
DB 170 ISCKAFSITGSSKEISFTNSTAKKGAIAATGIAHLSDNOGTIRFSGNAYVNSGAVYS 229

Db 531 YENHNKASSTYPLELTTAGANGTITLALSTLTQEPETHYGOAM-OLSMANATSS 589  
 QY 589 -TATFNWTKTGYIPNPERIGSLVPSNLMAAFIDISSLHMETANGLQDRAFCAGIS 647  
 Db 590 KGSIMWTITGIIIPSEKSNLPLNSLMGNFIDINSINLIETKSSGEPEBELMISGTA 649  
 QY 648 NEFHKDKTRKGFRLHSGGYIGGNLHATCSDKLISAPCOLFGDRDYFVAKNOGTVVG 707  
 Db 650 NEFYRDSMFTBGRFRIISGVALGITATPAEDQLTFACQFLPARNRNHITKKNEDTNG 709  
 QY 708 GILYOHNETTISLPCKL-----RPSLSYPTETIPVLPSGNLSTPTPDNDLTKYTTY 761  
 Db 710 ASLYFHETEGLEDIANFLNFKATRAPWVLSLSEIOLIPLSFDKFFSLYHDNMKTYTDTN 769  
 QY 762 PNYKSGWGDSDALEGGRAPICLDESALFEQYMPMKIQFYAAHOEGFEQGTAREEG 821  
 Db 770 SLIKSGWRNDACADGASLPFVIVSYPTLKEPEPVKQYIYAHQODYERABRAEN 829  
 QY 822 SSRVNLALPIGIRFDKESDCODATYNTLGYVDLVRSNPPCTTTLRISGDSMKTFTGN 881  
 Db 830 KSELINVEIPIGVTFERDSKSEKGYDTLMTYILDAYRNPQCQSLIASDANMMAVGTN 889  
 QY 882 LARQALVLRAGNHFCNSNFEPASQSFELRGSSRNRYNDLAKTIOF 928  
 Db 890 LARQGSVRANHFQVNPMEIFGQFAFEVRSSSRNRYNMLGSKFCF 936  
 RESULT 7  
 Q9RB66 PRELIMINARY; PRT: 930 AA.  
 AC 09RB66;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2001 (TREMBLrel. 16, Last annotation update)  
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE  
 DE PROTEIN G FAMILY).  
 GN PMP\_8 OR CP0307.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CJ38;  
 RX MEDLINE=2030349; PubMed=10871362;  
 RA Shirai T., Ishii K., Hattori M., Kubara S., Nakazawa T.,  
 RA Shiba T., Ishii K., Hattori M., Kubara S., Nakazawa T.,  
 RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,  
 RA McClarty G., Salberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL; AP002546; BAA98654.1; -;  
 DR EMBL; AE002193; AAF38164.1; -;  
 DR TIGR; CP0307; -;  
 DR InterPro; IPR003357; -;  
 DR InterPro; IPR003368; -;  
 DR Pfam; PF02385; OMP\_1; -;  
 DR Pfam; PF02415; DUF145; 1.  
 SQ SEQUENCE 930 AA; 97639 MW; 46A8896761391C09 CRC64;

Query Match 40.4%; Score 1929; DB 2; Length 930;  
 Best Local Similarity 43.4%; Pred. No. 1,5e-102;

Matches 414; Conservative 164; Mismatches 327; Indels 48; Gaps 17;  
 QY 1 MRSSEPKFVSFALEPISM-IAT---ETVLDSASFDGNKNQNSVRSODAGTYLFE 56  
 Db 1 MKIPLKLLISSTLVPIILSLTAITYGADSLSPDSEFDAGSGTTPKPKSTANGNINYL 60  
 QY 57 KGNVLTENIPGTATITKSCFNNTKGLTFNGNSLIEQYDAGTVAGAAVSSVYDS 116  
 Db 61 SGNVYI-NDAGGTLTGCCFFETTGDLTFQTKGYFSFNVTYDAGSNAGAA-STRADKA 118  
 QY 117 TTFIGSSLSFTASPOGSLTTGKAVSCSTGLKFDKNVSLFSKNEFD---NGCAITA 173  
 Db 119 LTFEGFNSLFTAAQGTIVASGKSTLS-SAGALNLDNCGTIIIFSQNVSEANNNGAITA 177  
 QY 174 KTLSTLGTMSALFESNTSSKKGAIGTSDALITNGOGEVSPDNTSSDSGAITEAS 233  
 Db 178 KTLSTISGNTSTFTFSNSAKKLGALYSSAASISGNTQOLVPMNKGEGGALGFENS 237  
 QY 234 VTISSNAKVSFLDNKVTGASSTTGDMSGALCAVYKSTDTKVTTLGNOMLFSNNTSTT 293  
 Db 238 SSTIQNSSLFFSGNATADAAG-----KGALYCEKTEGETPLTISGNKSLFPAENSVT 291  
 QY 294 AGGALYKRLLELASGGLTFSRNSVNGTAPKGAIAIEDSGELSLASDSGLIVFLGNTV 353  
 Db 292 OGGAICAHGLDLSAAGPTLFSNNRCGNTAAGKGAIAIADSGSLISANQGDITFLGNTL 351  
 QY 354 TSTTT-PGTRSSIDLTGSAKMTALRSAGRAIFYDPI---TTGSSTVTYDLKVNETPA 409  
 Db 352 TSTSAFTSTRNAIYLGSSAKITINLRAQOSIYTFDPLASNTTGA---DVLTIQNPDS 407  
 QY 410 DSALQYTGNIITFGEKLSLEADSKNLTSLKLPVTLSSGTLSLKHVTLQTAFTQA 469  
 Db 408 NSPLDVSGLTVFSGEKLSDAEKADNFTSILKOPALASGLTALGNELDVNGFTQNE 467  
 QY 470 DSRLENDVQTLIEPADTSTIN--NLVINSSIDGAKAKIERKANSKNTLSGTITLLDP 527  
 Db 468 GSTLLMQPTKULK-ADTEAISLTKLVVDLSALEGNKSVSIEFAGAKATTLVSPLEFQDS 526  
 QY 528 TGTFFENHSLRPOSYDIELKAS-----GYTSAVTPDPPIMGKFFHYGQGTWG 578  
 Db 527 SGNFYESHITINQAFIOPLVFPAATAASDIYIDALLTSFVQPEP-----HYGYGHE 580  
 QY 579 PLYVGTGASTATFPMNKTGYIPNPERIGSLVPSNLMAAFIDISSLHMETANGLQD 638  
 Db 581 AYVADTSTAKSGMTWVTGYNPERRASVYDDSLMASFTDIRLQIMTSQANSIYQO 640  
 QY 639 RAFWAGLSNFFHKDKSTRKRGFRHLSGTYVIGNHLTSDKILSAFQOLGRDRDYF 698  
 Db 641 KGLMASGTANFFHKDKSGTNOAFRRHSYGYIVGGSADSESENFSVAFQOLFCKDKDLFT 700  
 QY 699 AKNOGVYGGTLYVQHNETYISLPCKLRPC--SLSYVPTETIPVLESGNLSYHTDNDLKT 756  
 Db 701 VENTSINITYLASIXLRALGLGP---MPSFGSITMLADIPILNAOISYSTKNDMDT 757  
 QY 757 KTTTPTVYKSGWGDSDALEFGGRAPICL-DESALEEQYAMPKLOFYAAHOEGFEQGT 815  
 Db 758 RSTSYDEAGSWTNNNGALELGSLALYLPKRAPFOGYFPLKFOAVYSROONFKESGA 817  
 QY 816 EAREGSSRLVNLALPIGIRFDKESDCODATYNTLGYVDLVRSNPPCTTTLRISGDSW 875  
 Db 818 EARAEDDGLVNCISIPVGIIRLEKSIDENKNEISLAYIGDYVRKPNRRTSLMVGASW 877  
 QY 876 KTFEGTLARQALVLRAGNHFCNSNFEPASQSFELRGSSRNRYNDLAKTIOF 928  
 Db 878 TSLCKLARQALASAGSHLTLSPHELDSGEAAYELRGSAAHIYNDCCGLRYSE 930  
 RESULT 8  
 Q9Z393 PRELIMINARY; PRT: 930 AA.  
 AC 09Z393;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

QY 118 TFISSLSFIASPGSSIT-TGKGVSCSTSLKFDKRVSLFSKFNSTDNGAATLTKL 176  
 DB 121 LFNDSRLSTIISCSLSLSPGQCALK-SVGNLSITGNSQIITFONFSSDNGVINTNPF 179  
 QY 177 SLTGTTMSALPSEMT--SSKRGAIQTSALDALTIGNOGEVFSNDTSSDGAATFEASV 234  
 DB 180 LLSGTSOPASPSRNOAFPGKGGVYATGTTIENSFGIVFSQNLASGSGALYSTDNC 239  
 QY 235 TISNNAKVSFIDNKVYTGASSSTGTGDMGCAICAKTSTDVITVLGNOMLFSNNTSTA 294  
 DB 240 SITNFOVIFPGNSNMEWAQA----OGGAICCC--TTTDKVTYTLGNKNLSFTNTALTLY 292  
 QY 295 GGAIYVKKLELASGGLTLFSRNSVNGTAPK--GGAIAIEDSGELTSADSDIYFLGNT 352  
 DB 293 GGAISGLKVSISAGPITLFGSN-ISGSSAGOGGCAITASAGELALSATSQDITFNNQ 351  
 QY 353 VTSTPGTNRSSIDLGTSAKMTALSAAGRAIYFYDPIITGSSSTTVTDVLYKNETPADSA 412  
 DB 352 VTNGSTSR-RNAINIIDPAKVTISRAATGQSIYFYDPIITNGTAASTDTMLNLADANSE 410  
 QY 413 LQYGNIIITGEKISEFPADSKNLTSLKLPVTLSSGTLKKGVTYLTQTAFTQADSR 472  
 DB 411 IEYGAIVFSEKLSPEKATAANVTSTIRQPAVLARDGLVLRGQVTVTFKDLTQSPGSR 470  
 QY 473 LEMVGTLEPADTS-TINNVLINISSIDGAKKAKIETKATSKNLTLSGTTILDPTGTF 531  
 DB 471 ILMGGTTLKAKENLSINGLAVNLSSIDGINKALKTEADKNLSLSTGTTALIDTESF 530  
 QY 532 YENHSLRNPQSYDILELK--ASGTVTSTAVTPDPIMEKHFYOGTWGPIVWGTSAST 588  
 DB 531 YENHNLKASATYPLLELTITAGANGTITLGAISTLTLOEPETHYOGQNW-QLSMANATSS 589  
 QY 589 -TAFENWTGTYIPPEPISGLVPSNMANATIDISSLHYMETANEGLQGRATFCAGLS 647  
 DB 590 KIGSINMTRTYIIPSEKSNMLPLNSLMGNFIDIRSNOLLETSSSGEPFRELMLSGIA 649  
 QY 648 NEFKDSTKTRGFRHLGSGVIGGNLHTSCDKILSAFCQLFGDRRYFAKNGCTYAG 707  
 DB 650 NEFRDMSPTNHRHRSISGAYALGTTATTPAEDOLTFACQLFARDNRHITGKNNGDITG 709  
 QY 708 GTLYXHNETHYISLPCKL-----RPSLSIYVPEIPIVLFSGNLSYTHTDNLTKYTTY 761  
 DB 710 ASLTFHTEGLFDIANFLMGKATRAPVWLSISQIIPLSFAKKSYLTDHMKMTYYIDN 769  
 QY 762 PTYVSGNDSFALEFGRAPICTDESALFEQYMPKLGQVYVAHQSEFGQTEARFEG 821  
 DB 770 SIINGSWRNDAFCADLGLSLEFVLSVPLAKEVEPFVVOYIYAHQODFYERHAEGRAFN 829  
 QY 822 SSRVLNALPIGIRFDESDCDATYNLTGLYVDLVRSNPDCTTTLRISGDSMKTFTGN 881  
 DB 830 KSELINVEIPIGVFERBRSKSEKGTIDTLIMYIIDAYRANKCOTSLASANMAAYGTN 889  
 QY 882 LARQALVLRAGNHCFNSNFEAFSOFSEFLRSGSSRNTNVLDGARYOF 928  
 DB 890 LARGFSVRANHFQVNPHEITFGQAFEVYSSRNTNLTGSKFCF 936

RESULT 6  
 Q9J542 PRELIMINARY: PRT: 936 AA.  
 AC Q9J542: 01-OCT-2000 (TEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)  
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).  
 GN PMP.7 OR CP0308.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiaophila.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-J138; PubMed:10871362;  
 RX Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CWL029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Reed T.D., Brunham R.C., Shen C., Gall S.R., Heideberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umeyam L.A., Utterback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,  
 RA McClarity G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39."  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AP002546; BAA98653.1; -;  
 DR EMBL: AE002193; AAP38165.1; -;  
 DR TIGR: CP0308; -;  
 DR InterPro: IPR002637; -;  
 DR InterPro: IPR003357; -;  
 DR InterPro: IPR003368; -;  
 DR Pfam: PF02385; OMP.1.  
 DR Pfam: PF02415; DUF145; 1.  
 DR Pfam: PD004952; -; 1.  
 SQ SEQUENCE 936 AA; 100105 MW; 3981DB3C950AF95A CRC64;

Query Match 40.8%; Score 1950; DB 2; Length 936;  
 Best Local Similarity 43.1%; Pred. No. 9.7e-104;  
 Matches 408; Conservative 179; Mismatches 330; Indels 30; Gaps 16;

QY 1 MKSSPEKFEVSTFAIF-PLSMIAETVLDSS-ASEDGNKNGNFSVRESQE-DAGTLYLEK 57  
 DB 1 MKSSWMLFFSISPIPLFSSLSIAAEVLTDSSNNSYDGSNGTTFYFVSTDAAGTYSLL 60  
 QY 58 GNVLTENIPGIGTALTKCFNNTKGDLTFTGNGNSLFEQYDAGTVAGAAVNSVVDKST 117  
 DB 61 SDVSFONNGALGIPLASCFLFAGDGLTFQGNHAKLFAFNAGSSACTVASTAADNLL 120  
 QY 118 TFISSLSFIASPGSSIT-TGKGVSCSTSLKFDKRVSLFSKFNSTDNGAATLTKL 176  
 DB 121 LFNDSRLSTIISCSLSLSPGQCALK-SVGNLSITGNSQIITFONFSSDNGVINTNPF 179  
 QY 177 SLTGTTMSALPSEMT--SSKRGAIQTSALDALTIGNOGEVFSNDTSSDGAATFEASV 234  
 DB 180 LLSGTSOPASPSRNOAFPGKGGVYATGTTIENSFGIVFSQNLASGSGALYSTDNC 239  
 QY 235 TISNNAKVSFIDNKVYTGASSSTGTGDMGCAICAKTSTDVITVLGNOMLFSNNTSTA 294  
 DB 240 SITNFOVIFPGNSNMEWAQA----OGGAICCC--TTTDKVTYTLGNKNLSFTNTALTLY 292  
 QY 295 GGAIYVKKLELASGGLTLFSRNSVNGTAPK--GGAIAIEDSGELTSADSDIYFLGNT 352  
 DB 293 GGAISGLKVSISAGPITLFGSN-ISGSSAGOGGCAITASAGELALSATSQDITFNNQ 351  
 QY 353 VTSTPGTNRSSIDLGTSAKMTALSAAGRAIYFYDPIITGSSSTTVTDVLYKNETPADSA 412  
 DB 352 VTNGSTSR-RNAINIIDPAKVTISRAATGQSIYFYDPIITNGTAASTDTMLNLADANSE 410  
 QY 413 LQYGNIIITGEKISEFPADSKNLTSLKLPVTLSSGTLKKGVTYLTQTAFTQADSR 472  
 DB 411 IEYGAIVFSEKLSPEKATAANVTSTIRQPAVLARDGLVLRGQVTVTFKDLTQSPGSR 470  
 QY 473 LEMVGTLEPADTS-TINNVLINISSIDGAKKAKIETKATSKNLTLSGTTILDPTGTF 531  
 DB 471 ILMGGTTLKAKENLSINGLAVNLSSIDGINKALKTEADKNLSLSTGTTALIDTESF 530  
 QY 532 YENHSLRNPQSYDILELK--ASGTVTSTAVTPDPIMEKHFYOGTWGPIVWGTSAST 588

of autotransporting pathogenicity factors.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE-20330349; PubMed-10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishi K., Hattori M., Kohara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CML029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE-20150255; PubMed-10684935;  
 RA Reed T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Ulteback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL; AE001628; AAD18591.1; -;  
 DR EMBL; AJ133034; CAB37069.1; -;  
 DR EMBL; AP002546; BAA38655.1; -;  
 DR EMBL; AE002192; AAF38163.1; -;  
 DR TIGR; CP0306; -;  
 DR InterPro; IPR003357; -;  
 DR InterPro; IPR003368; -;  
 DR Pfam; PF02385; OMP; 1.  
 DR Pfam; PF02415; DUF145; 1.  
 KW Signal.  
 FT CHAIN 1 27 POTENTIAL.  
 FT SEQUENCE 928 AA; 98332 MW; 58910A8F04F1219 CRC64;  
 Query Match 41.4%; Score 1977; DB 2; Length 928;  
 Best Local Similarity 45.1%; Pred. No. 2,7e-105;  
 Matches 428; Conservative 149; Mismatches 329; Indels 44; Gaps 19;  
 QY 1 MKSSPEKVEYSPFAIFPLSM-----IATEVLDSSASFDGKNGNFVRESQEDA-GTT 53  
 DB 1 MKSSLMFELISSSLALPLSLINFSAPFAVEINLGPNTSGS--PGHYTPPAQTNNADGTI 58  
 QY 54 YLEKGNVTELEIPGCTATTKSCFNTKGDLTFTGNGNLLTQYVDAGVGAANVSSV 113  
 DB 59 YMLTGVSTLN-AGSEPTALTASCPEKETTGNLSFGHGVOFLQNTDAG--ANCTFTNTWA 115  
 QY 114 DKSTFIFGSSLSFTASPCSSITTTGKAVSCVSGSLKPKNVSLLSFKNFSTDNGAITA 173  
 DB 116 NKLLSSGSGSYLSLQTF--TNMTTGAIK-STGACSIOSNISCYFGQNFSDNGALOG 172  
 QY 174 KTLSTGTMTSALFSEMTSSKKGAIQTSDALTTGNOGEVSPSDNTSSDGAALFTFAS 233  
 DB 173 SSSISLS-LNPNTLTFKAKTKKATKCGALYSTGITTINLTLSASFSEMTAANNGAITYEAS 231  
 QY 234 VTISSNAKXSEFDNTVGTGSSSTGTGMSGATCAKTSPTDTVY-TLTGOMLLFSNNTST 292  
 DB 232 SFTSSKKAISFTINNSTATSA-----TGAALYCSSTSAKPVLTLSDNGELNFTGNTAI 285  
 QY 293 TAGGATVYKLELAGGLTLFGRNSVNGCTAPKAGAIATEDSGELSLSDSGDIYFLGNT 352  
 DB 286 TSGGATYTNVLVSSGGLTLFKNNSAIDRAPLGALAIADSGSLSLAGDITFEGMT 345  
 QY 353 V--TSTPTGTRSSIDLG-TSAKMTALRSAAAGRAIYFDPTITGSSSTVTVDLAKNETP 408  
 DB 346 VYKASSSSQTTTRNSINIGNTNAKIYQLRASSQNTIYFDYPTITSTITALSLDALNNGPD 405  
 QY 409 ADSALQYTNITFTGKLESETEAOSKNTLSKLTGQVNTLSGGTSLSKHVTLOTOTAFQ 468  
 DB 406 LAGNPAYQGTIVFSGEKLSEARAEADNLKSTIOQPLTLIAGGOLSLKSGVTLVAASFSS 465

ADSRLEMDGTTLEPADSTINNVLINISSIDGAKKAKIETKATSKNLTLSCTITLDDPT 528  
 DB 466 PESTLMDAGTTLEPADGTINNVLVNDLSLETKKATLAKATQATYVTLSSSLVSDS 525  
 QY 529 GTFENHSLRNPQSYDILELAKS--GVTSYAVYDPPIMGEKFRHYGQGTGPYWGTC 586  
 DB 526 GNVEDVSNWNPQVFSCLTTLTADDDPANHITDLADPLEKNPIHNGYQGNMA-LSMOEDT 584  
 QY 587 ST---TATFNMTKGYIPPERISGLVPNSLNAPFIDISSLIYMETANEGLOGRAFVC 643  
 DB 585 AKSKRAATLTWTKTGNPBERGTLVANTLWGSFVDVNSIOQLVATKROSQETRTGWC 644  
 QY 644 AGLSNFHKDSTKTRGFRHLISGCVYVIGNLHTCSDKILSAFCOLLFGRDRPYFAKNQ 703  
 DB 645 EGISNPFHKDSTKINKKPFHISAGYVAGTTTLASGNLITAFCOLFGDRDRHFINKNA 704  
 QY 704 TVYGGTLYYQHNEYIYSLPCRLPQSLSTVP--TEIPVLSFGNLSYHTTDLTKTYT 760  
 DB 705 SAYAASLHQHLATLSS-----PSILRLTPGSESEQPLFPAQISYIYSKNTMKTYTQ 758  
 QY 761 YPTVAGWGNDSFALFEGGRAP-ICLDESALFEQYMPEMKLOFYVAHOBFKEQGT 818  
 DB 759 APRGSSWYNDGCALELASSLPHALSHGELHAFPEFTKVASYIHODSEKERTTLVR 818  
 QY 819 EFGSSRLVNLALPIGIFPKESDCODATYNTLTGYTVDLVRSNPQCTTLRTISGSMKTF 878  
 DB 819 SFDSDGLINVSYPIDGTFPRFRSNERASYEATVIYADVRRKNPDCTTLTLNNSMKKT 878  
 QY 879 GTNLARQALVLRAGNHCFNSNFEAFSOPFSELRGSSRNRYNDLAKAYOF 928  
 DB 879 GTNLSKQAGIGRAGIFAFSPNLEVTSLNLSMEIRGSSRSRYNADLGKGF 928  
 RESULT 5  
 Q92898 PRELIMINARY; PRT: 936 AA.  
 AC Q92898;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY.  
 GN PMP\_7.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CWL029;  
 RX MEDLINE-99206606; PubMed-10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammell C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 DR EMBL; AE001627; AAD18589.1; -;  
 DR InterPro; IPR003357; -;  
 DR InterPro; IPR003368; -;  
 DR Pfam; PF02385; OMP; 1.  
 DR Pfam; PF02415; DUF145; 1.  
 SO SEQUENCE 936 AA; 100079 MW; 8881D78A53D194EC CRC64;  
 Query Match 40.8%; Score 1951; DB 2; Length 936;  
 Best Local Similarity 43.1%; Pred. No. 8,5e-104;  
 Matches 408; Conservative 179; Mismatches 330; Indels 30; Gaps 16;  
 QY 1 MKSSPEKVEYSPFAIF-PLSMATEVLDSS-ASFDGKNGNFVRESQE-DAGTYTLK 57  
 DB 1 MKSSVWLFESSIPFSSSIIVAAEVTLDSSNNYSYDGSGETTFVSTDAAGTYSL 60  
 QY 58 GVNTLEINGTGTATKSCFNTKGDLTFTGNGNLSLFTQVAGVGAANVSSVYDKST 117  
 DB 61 SDVSRONAGALGIPLAGCGFLRAGDITFGQNHAKLFAFINAGSAGIVASTAADK 120

```

Db 643 WAAGVAFNLDKDKKGRKRYRHKSGGYAIGGAOQTCSENLISFAFCOLLFGSKDDELVAKN 702
QY 702 QGVYGGTLYVQHNEITYISLPCRLPCSLAYVP---TEIPVLFSGNLSYTHTDNDLTKY 758
Db 703 HTDYYAGAFYIQH---ITECSGFICGLDLPKPSWSHKRPVLEGOALYSHVSDNLDKTY 758
QY 759 TTYPTVAGSGWGNDSFALFEFGGRAPICLDESALFEQYMPFMKLOFVYAHQEFKRGTEAR 818
Db 759 TAYEVGSGWGNNAFNMMLGASSHSYPEYLHCFTPYAPYIKLNLTYIRQDSFSEKTEGR 818
QY 819 EFGSSRLVNLALPIGIRFDEKSDCODATYNTLTGYVDLVRSNPDCTTLTISGDSWKT 878
Db 819 SFDDSNLFNLSLPIGVAFKESDCNDPSYDLTLTSYVDLINDPKCTTALYISGASMETY 878
QY 879 GTNLARQALVLRAGNHPCFNSFEAFSOFSELRGSSRNRYVDGAKYQF 928
Db 879 ANNLRQALQVRAGSHYAFSPMEFVLQGFVEVGRSSRIYVDLGGKQF 928

RESULT 3
086163 PRELIMINARY; PRT; 914 AA.
AC 086163;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE OUTER MEMBRANE PROTEIN 5 (FRAGMENT).
GN OMP5.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC/CWL-029/VR-1310;
RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RT Identification of two novel genes encoding outer membrane complex
RL associated surface layer proteins in Chlamydia pneumoniae.*;
EMBL: A001311; CA04671.1; to the EMBL/GenBank/DBJ databases.
DR InterPro: IPR003357; -.
DR InterPro: IPR003366; -.
DR Pfam: PF02385; OMP; 1.
DR Pfam: PF02415; DUF145; 1.
KM Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 >914 OUTER MEMBRANE PROTEIN 5.
FT NON_TER 914 914
SQ SEQUENCE 914 AA; 95603 MW; 8BF33BAB680F5E3 CRC64;

Query Match 42.1%; Score 2011; DB 2; Length 914;
Best Local Similarity 46.5%; Pred. No. 3e-107;
Matches 433; Conservative 140; Mismatches 317; Indels 44; Gaps 20;

```

```

QY 291 STTAGAIYVKLELAS--GGTLFERNVNGTAPKGAIAIEDSGELSDGDIYF 348
Db 284 AVANGAIIYAKKLLASGGGGISFNINVOGTAGNGAIIAAGCSIASAGDITF 343
QY 349 LGNTVSTTP-GNRRSIDLGTSAKMTALRSAGRAIYFDPTTGSSTVTVLKYNET 407
Db 344 NGNLIYATPQTNRNSIDISTAKITNLRAISGHISFEFYPITANPADSDTLNLNKA 403
QY 408 PADBALQYTGNIITFGEKLSFTEAADSKNLTKLQPTVYLSGGLSLKHGVTLOQAFTQ 467
Db 404 DAGNSTYSGSIVSGEKLSEDEAKVADNLSTLQKPTVLAGMLYKRGVTLDTKGTQ 463
QY 468 QADSRLEMDVGTLEPA-DTSTINNLVINISSIGAKKATETKATSKNLTSGITFLD 526
Db 464 TAGSSVIMADGTTLKASIEEYTLGLSLPVDLSDEGKKVYVLAASAKNVALSGPILLD 523
QY 527 PTGTFEYENHSLRNPQSDYLELKASGVTVSTAVTPDPIMGEKPHYGYOGTWGPVW--- 582
Db 524 NQGNAYERNHDLGKTQDSFVOLSAIGTATTPVPAVPTVATPRTHYGQGTWG-MTWVDDT 582
QY 583 -GTGASTTATPTNMTKGTIIPPERIGSLVPRNSLNNAFIDISLSHYMETANEGIQDRA 641
Db 583 ASTPKTKTATLAWNTGTGLPNERQGLVPNSLWGSFSDIQAIGVIERSLATLCSDRGF 642
QY 642 WCAGLSNFHNDSTKTRGFPHLSGVYIGNLHTCSDPKILISAFCOLFGDRDPYFAKN 701
Db 643 WAAGVAFNLDKDKKGRKRYRHKSGGYAIGGAOQTCSENLISFAFCOLLFGSKDDELVAKN 702
QY 702 QGVYGGTLYVQHNEITYISLPCRLPCSLAYVP---TEIPVLFSGNLSYTHTDNDLTKY 758
Db 703 HTDYYAGAFYIQH---ITECSGFICGLDLPKPSWSHKRPVLEGOALYSHVSDNLDKTY 758
QY 759 TTYPTVAGSGWGNDSFALFEFGGRAPICLDESALFEQYMPFMKLOFVYAHQEFKRGTEAR 818
Db 759 TAYEVGSGWGNNAFNMMLGASSHSYPEYLHCFTPYAPYIKLNLTYIRQDSFSEKTEGR 818
QY 819 EFGSSRLVNLALPIGIRFDEKSDCODATYNTLTGYVDLVRSNPDCTTLTISGDSWKT 878
Db 819 SFDDSNLFNLSLPIGVAFKESDCNDPSYDLTLTSYVDLINDPKCTTALYISGASMETY 878
QY 879 GTNLARQALVLRAGNHPCFNSFEAFSOFSELRGSSRNRYVDGAKYQF 914
Db 879 ANNLRQALQVRAGSHYAFSPMEFVLQGFVEVGRSSRIYVDLGGKQF 914

RESULT 4
092398 PRELIMINARY; PRT; 928 AA.
AC 092398;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE OUTER MEMBRANE PROTEIN PRECURSOR (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
GN OMP10 OR PMP 9 OR CP0306.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.*";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310;
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birkelund S.;
RT "Chlamydia proteins containing the GAT-repeat belong to a subfamily

```

```

DB 121 TLTGNOMLFFSNNTSTTAGAIVYKLIELASGGLTFFSNVNGGAPKGAIAIEDSGE 180
QY 337 LSLSDSGDIVLGNVSTTPTGNTSSIDLGTSAKMTLRSAAGAIYFDPITGSSGT 396
DB 181 LSLSDSGDIVLGNVSTTPTGNTSSIDLGTSAKMTLRSAAGAIYFDPITGSSGT 240
QY 397 TTTVDLVKVEFPADSAALQYTGNIIFTEGKLSFTEADSKNLSKLLQPTLSGGLSLKH 456
DB 241 TTTVDLVKVEFPADSAALQYTGNIIFTEGKLSFTEADSKNLSKLLQPTLSGGLSLKH 300
QY 457 GVTLOQAFTOADSRLEMDVGTLEPADTSTINNLVINISSIDAKKAKIETKATSKNL 516
DB 301 GVTLOQAFTOADSRLEMDVGTLEPADTSTINNLVINISSIDAKKAKIETKATSKNL 360
QY 517 TLTSGTTLTDPGTETENSLNPOSYDILELKASCTVSTAVTPDPIKGEKPHYQGT 576
DB 361 TLTSGTTLTDPGTETENSLNPOSYDILELKASCTVSTAVTPDPIKGEKPHYQGT 420
QY 577 WGPVWGTAFTAEENMTKGTGYPNPERIGSLVPSLNNAFIDISLHYMETANEGLQ 636
DB 421 WGPVWGTAFTAEENMTKGTGYPNPERIGSLVPSLNNAFIDISLHYMETANEGLQ 480
QY 637 GDRAPWCAGLSNFEHKSRTKRRGFRHLSGGYVIGGNLHTCSDKILSAFCOLFGDRDY 696
DB 481 GDRAPWCAGLSNFEHKSRTKRRGFRHLSGGYVIGGNLHTCSDKILSAFCOLFGDRDY 540
QY 697 FVAKNGTGYGGGLYQHNETYISLPCRLPCSLSVPEIPVPSGNLSYTHTDMLT 756
DB 541 FVAKNGTGYGGGLYQHNETYISLPCRLPCSLSVPEIPVPSGNLSYTHTDMLT 600
QY 757 KYTTPYTVKSGWGNDSFALFEGGRAPICLDESALFEQYMPFKLOFVYAHQEGFKQGT 816
DB 601 KYTTPYTVKSGWGNDSFALFEGGRAPICLDESALFEQYMPFKLOFVYAHQEGFKQGT 660
QY 817 ARFSGSRVYNALPGIRFDKSDCQATYNTLGYTYVDVLRNSPDCITTLISDSMK 876
DB 661 ARFSGSRVYNALPGIRFDKSDCQATYNTLGYTYVDVLRNSPDCITTLISDSMK 720
QY 877 TGTCTNARQALVLRAGNHCFNSNFEAFQSFELGSSRNYNVDGAKYOF 928
DB 721 TGTCTNARQALVLRAGNHCFNSNFEAFQSFELGSSRNYNVDGAKYOF 772

RESULT 2
Q9RB65 PRELIMINARY: PRT: 928 AA.
AC Q9RB65: Q9RB64: Q9SBP2: 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (OUTER MEMBRANE PROTEIN 5
DE PRECURSOR (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
DE PMP_10 OR OMP5 OR CP0303.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shira M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kohara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CHL029 from USA.";
RL Nucleic Acids Res. 28: 2311-2314 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310;
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birkegaard S.;
RT "Chlamydia proteins containing the GGA1-repeat belong to a subfamily
RT of autotransporting pathogenicity factors.";
```

```

RL Submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Utechtack T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gilm M., Nelson W., DeBoy R., Kolonay J.,
RA McCarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28: 1397-1406 (2000).
DR EMBL: AP002546; BAA98657.1; -
DR EMBL: AJ133034; CAB37071.1; -
DR EMBL: AE002192; AAF38160.1; -
DR TIGR: CP0303; -
DR InterPro: IPR003357; -
DR Pfam: PF02385; OMP; 1.
KW Signal.
SQ SEQUENCE 928 AA; 97229 MW; 0590D5206A1DD0E1 CRC64;

Query Match 43.3%; Score 2069; DB 2; Length 928;
Best Local Similarity 46.9%; Pred. No. 1.5e-110;
Matches 446; Conservative 141; Mismatches 319; Indels 44; Gaps 20;

QY 1 MKSSPFKEVF-STFAIF---PLSMITEVTVLDSASFQDNKN-GNFSVESQDAGTYL 55
DB 1 MKSOFMWLVLSSTLACFTSCSTYFAATAENIGPSDFSSTNGITYTPKNT--TTGIDYT 58
QY 56 FKNWTLNENIPGTGTAITKSCFNNTKGLFTFGNGNSLLFQVYDAGTVAGAAVNSVDK 115
DB 59 LQGDITLQNL-GDSALTLGCGFSDTETESLSPAGKGYSLFLNKS-SARGAAL-SVYTK 115
QY 116 STTFIFGSSLSFLASGSSITF--GKAVSCSTGSLKFPKNNLSLFSKPFSDNGAITA 173
DB 116 NLSLTFGSSLSFLAAPSIVTTPSGGAVKCG-GDLTFNNGITLTKODYCENGAISTF 174
QY 174 KTLISLGTMSALFSENTSS---KKGAIOISDALITGNOGEVSESDMTSSDGAIFT 230
DB 175 KTLISLKNSTGSISEFGNKSATGKKGALCATCTVITNTNAPTLFSNNIAEAGAIINS 234
QY 231 EASVTSNNNAKVSFIDNKTGASSSTTGDMSGAICAYTSTDTKVTLTGNOMLFSNNT 290
DB 235 TGNCITGTNTSLFSENSVT---ATAG--NGAL-----SGADVTISGNSVTFSGNO 283
QY 291 STTAGCAIYVKLIELAS--GGLTFEKRNSVNGTAPKGAIAIEDSGEISLSDSDIYF 348
DB 284 AVANGCAIYAKKLIELASGGGGGSGISFNNIVOGTTAONGAISTLAAGECSLSAEGDITF 343
QY 349 LGNTVSTTP-GNRRSIDLGTSAKMTLRSAAGAIYFDPITGSSRNYNVDGAKYOF 407
DB 344 NGNAIVATTPPTTKRNSIDIGSTAKTINRAISGHSIFEDPTTANWADSDTDLNLNKA 403
QY 408 PADSALOYTGNIIFTEGKLSFTEADSKNLSKLLQPTLSGGLSLKHGVTLOQAFTO 467
DB 404 DAGNSTDYGSGIVFSEKLSDEDEAKVADNLSTLKGVLTPVLTGNGVLKGVLTDTGTGFP 463
QY 468 QADSRLEMDVGTLEPA-DTSTINNLVINISSIDAKKAKIETKATSKNLTSGLTTLTD 526
DB 464 TAGSSVIMAGCTTLKASFEVLTGLISIPVDSIGEEKKVVIASAASKVVALSGPLILD 523
QY 527 PGTETENSLNPOSYDILELKASCTVSTAVTPDPIKGEKPHYQGTGWPPIVW---- 582
DB 524 NQGNALYENDLGTQDFSVQSLASTATTTDPAVPTVATPHHYQGTGW-MTWVYDT 582
QY 583 -GTGASTATFMTKGTGYPNPERIGSLVPSLNNAFIDISLHYMETANEGLQDRAAF 641
DB 583 ASHPKKTATLMTNTGYLPNPERQGPLVPSNLMGFSFDIQAQVIERSAITLCSDRG 642
QY 642 WCAGLSNFEHKSRTKRRGFRHLSGGYVIGGNLHTCSDKILSAFCOLFGDRDYVAKN 701
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 2, 2001, 03:31:23 ; Search time 119.05 Seconds  
(Without alignments)  
1031.324 Million cell updates/sec

Title: US-09-446-677B-10  
Perfect score: 4778  
Sequence: 1 MKSFFPKVFSTFAIFPLSM.....FELRGSSRNVDLAKYQF 928

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_16:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_unclassified:\*  
13: sp\_vertebrate:\*  
14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3991	83.5	772	2	Q9RB71 chlamydia p
2	2069	43.3	928	2	Q9RB65 chlamydia p
3	2011	42.1	914	2	Q9RB63 chlamydia p
4	1977	41.4	928	2	Q9RB38 chlamydia p
5	1951	40.8	936	2	Q9RB98 chlamydia p
6	1950	40.4	936	2	Q9RB66 chlamydia p
7	1929	40.4	930	2	Q9RB66 chlamydia p
8	1925	40.3	930	2	Q9RB66 chlamydia p
9	1908	39.9	926	2	Q9RB66 chlamydia p
10	1853	38.8	928	2	Q9RB66 chlamydia p
11	1853	38.8	928	2	Q9RB66 chlamydia p
12	1650.5	34.5	839	2	Q9RB29 chlamydia p
13	1637.5	34.3	847	2	Q9RB29 chlamydia p
14	1598.5	33.5	846	2	Q9RB29 chlamydia p
15	1562	32.7	841	2	Q9RB29 chlamydia p
16	1442	30.2	1276	2	Q9RB29 chlamydia p
17	1435.5	30.0	922	2	Q9RB29 chlamydia p
18	1434.5	30.0	922	2	Q9RB29 chlamydia p
19	1423.5	29.8	922	2	Q9RB29 chlamydia p

20	1381.5	28.9	1407	2	Q9RB99 chlamydia p
21	1377.5	28.8	973	2	Q9RB96 chlamydia p
22	1377.5	28.8	995	2	Q9RB96 chlamydia p
23	1235	25.8	712	2	Q9RB73 chlamydia p
24	1115.5	23.3	1013	2	Q9RB73 chlamydia p
25	1053	22.0	987	2	Q9RB73 chlamydia p
26	1017.5	21.3	445	2	Q9RB73 chlamydia p
27	935.5	19.6	649	2	Q9RB73 chlamydia p
28	916	18.2	359	2	Q9RB73 chlamydia p
29	905	18.9	867	2	Q9RB73 chlamydia p
30	854	17.9	878	2	Q9RB73 chlamydia p
31	819	17.1	494	2	Q9RB73 chlamydia p
32	801	16.8	427	2	Q9RB73 chlamydia p
33	754	15.8	186	2	Q9RB73 chlamydia p
34	680.5	14.2	1609	2	Q9RB73 chlamydia p
35	680.5	14.2	1609	2	Q9RB73 chlamydia p
36	672.5	14.1	978	2	Q9RB73 chlamydia p
37	670	14.0	947	2	Q9RB73 chlamydia p
38	667.5	14.0	978	2	Q9RB73 chlamydia p
39	666	13.9	947	2	Q9RB73 chlamydia p
40	664	13.9	1723	2	Q9RB73 chlamydia p
41	664	13.9	1723	2	Q9RB73 chlamydia p
42	664	13.9	1732	2	Q9RB73 chlamydia p
43	663.5	13.9	946	2	Q9RB73 chlamydia p
44	663.5	13.9	946	2	Q9RB73 chlamydia p
45	642	13.4	952	2	Q9RB73 chlamydia p

## ALIGNMENTS

RESULT 1  
ID Q9RB71 PRELIMINARY; PRT: 772 AA.  
AC Q9RB71;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE PMP-3.  
GN PMP-3.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SRRATN-J138.  
RX MEDLINE=2030349; PubMed=10871362;  
RA Shiba T., Ishii K., Hattori M., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CML029 from USA".  
RL Nucleic Acids Res. 28:2311-2314(2000).  
DR EMBL: AP002545; BAA98226.1; -.  
DR InterPro: IPR003357; -.  
DR Pfam: PF02385; OMP. 1.  
DR Pfam: PF02415; DUF145. 1.  
SQ SEQUENCE 772 AA; 82931 MW; 484FC56D635801EB CRC64;

Query Match 83.5%; Score 3991; DB 2; Length 772;  
Best local Similarity 99.9%; Pred. No. 9, 5e-221;  
Matches 771; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 157 LFSKNSSTNGCAITAKTSLTGTTMSALFSESTSSKKGCAIOTSDALITGNGGVSVF 216  
Db :|||||  
QY 1 MFSKNSSTNGCAITAKTSLTGTTMSALFSESTSSKKGCAIOTSDALITGNGGVSVF 60  
Db :|||||  
QY 217 SDNTSSSGAIFTEASVTSSNNAKVSFIQNKVTGASSTTGMSGCAICATYTSPTKV 276  
Db :|||||  
QY 61 SDNTSSSGAIFTEASVTSSNNAKVSFIQNKVTGASSTTGMSGCAICATYTSPTKV 120  
QY 277 TLGNGQLFSSNNSTTAGCAIYVKKLELASGILTFSSRSNVNGCTAPKCAIATIEDSGE 336



```

OY 1725 GAAGCTCACTGATGATATGTAAGGNAACATTTATGAAATCATATGTCAGCCATGA 1784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25628 GCGTATCATGCTGTGATAGTGCACAGGGAACTATTTATGCACTCATATCTGACACAA 25667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1785 CCAGCTCTTCTCTATTAATAAATCAAGGTTGATGCTGATGTTGATTAAGCTTACAT 1844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25688 GGAGGCTCTTCTTAATAGACCTTTCTGCACAGGAAGACGATGACTACAGATATCC 25747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1845 CAGCAGCCTTATCCCTGTTCCCTGCTGAGGATCTTAATCAGATACGATTCACAGGACA 1904
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25748 GCATACCCCAATTCATAAATACATCAATCATAT-----GGGATATCAAGGAAAC 25796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1905 ATGGAATGTTAATGAGTACGATACGATACGATCAATAATACAAAAGAGCCAGCAACTGG 1964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25797 TGGAAATATGTTTGGGTGAGATGACATGCAACTGCAAAAACAAAATGCTACTTAACTTG 25856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1965 GACCAAAACAGATTTGTTCCAGCCCGAAAGAAAATCTGCGTTAGTATGCAATACCT 2024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25857 GACTAAACAGGATACAAAGCCGAATCCAGAACGTCAGGACCTTGGGTCTAATAGCT 25916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2025 ATGGGAGCTTACTGACATCTGCTCTGACACAGCTTGTAGAGATCGGCAACTGG 2084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25917 GTGGGGTCTTTGTGATGATCGCTCCATTCAGAGCTCATGAGCCGAGCAAGTTG 25976
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2085 TATGNAACACAACAGGTTCTGGGTTCTCTCATGACGAACTTCTGATACACTGG 2144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25977 GTTATCTTGTACAAATTTGTGGGTATCAGGAATCGGACTTTTGTGATGAAGATCA 26036
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2145 AGATGAAAATCGCAAAAGCTTCGTCATACCTCTGAGAGCTACGTATCGTGAAGTGC 2204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26037 GAAAGGAACCAACGTATGATTCATTCATTCAGCGGGGTATGCAATGAGAGAGAT 26096
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2205 TCACACTCTTAAGACGACTTTTACCTTTCGCTTCGCTTCCTTCTGCTAGAGACAA 2264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26097 CTTCACGGCTCTGAAAATTTCTTAATTTGCTTTTGTCTGACCTTTTGGCTACGACAA 26156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2265 AGATGTTTATCTGCTACACACACTCTGACACTACGGTGAACCTTTATCTTCAAGCA 2324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26157 GGACCATCTTGTGGGTAGAACACATCCATGATATGAGGGGCAATGACATTACGACA 26216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2325 CTTCATACCTCTACCAACCCAAACTATTTGAGATTAGAGAGCAAAAGTTTCTGAATC 2384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26217 C-----CTCGAGAGTCTAAGACCCCTGCTAA 26243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2385 AGCTATGAAAAAATTCCTTAGGAAATTCCTTACCTTGATGTCGAATTCGTTG 2444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26244 GATTTGTGAGAAATTCGACTCCCTACCTTTGTCTTCAATGCTCGGTTGCTTATGG 26303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2445 CCATTTCAGACACCTGATGAAAGCACTATACCTATTCAGAGATCCGAAGTTCTTG 2504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26304 CCATACCGACATATACATGACCAACAAAGTACACTGGCTATTTCTCTGTTAAGGAAAGCTG 26363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2505 GAGCAACGAGTATAGCTGTGTGAGGCTAGACCTTCCTTTGTTCTTTCACACC 2564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26364 GGGAAATGATGCTTCGATGAAATGAGAGACCTATCCGATGATGCTTCAAGGACG 26423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2565 ACATCTCTTTTCAAGACCTTATTCACAGATGAAAGCGAAATGTTTATGATACACA 2624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26424 TGGGTCTTGGGTGATACCCACACGCACTTTTAAACCTAGAGATGATCTATGACATCA 26483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2625 AATATGCTTCTTGAAGCTCTAGTATGAGCCGTTTATGATTTAGTATGAAAGCTGCTTAA 2684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26484 GATATGCTTTAAGGAAAGGAGCAAGAGGCGCTTCTTCCAAAGTGAAGACTCTTCA 26543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2685 CCTCTGATCTCTGTGGGTGCAAAATTCGTGCAAGGGGATATGAGATTCCTACACCTA 2744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26544 TCTAGGGTCTCTAGGATTAATTT-----TGAGAAATTCCTCGATTAAGTCTACGTA 26597
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2745 TGATCTCTAGGATTTCTTGTTCGATGCTATGCTATACAAATCCCAATCTACAGCGAC 2804
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26598 TGAATCTCTCAATAGCTTACGTTCCGATGTGATGTGATCAAGGCTGACGACAAAC 26657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

OY 2805 TCTTGATGAGCCAGACTCTTGGAAAATTCGCGGTGGCAATCTTTCAGACAGCAAT 2864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26658 TCTTATGTTTCTGGGGATTTCTTGGTCAGACATGTGTAACAGCTTGTCTAGACAAAGCTCT 26717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2865 TTTACTGAGGGGTACCAACACTACGCTACACACTGCAATTTGTAGCTCTTGGACATTA 2924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26718 TCTTGATGCTGTGAAATTCATCATGCTTGTCTTCAAACTTGAAGTTTTCAGTCAGTT 26777
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2925 CCGTATGACCTCCGATCTTCAAGGACATCAATGATGTTGGTACCAACTCCG 2984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26778 TGAAGTGGATTGCCAGGGTCTTCTGTAAGTATGCTATGATCTTGGAGAAATTCGG 26837
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2985 ATTCTA 2990
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26838 ATTTTA 26843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: October 2, 2001, 00:28:56  
Job time: 42409 sec



KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
SOURCE

Chlamydia pneumoniae J138 (strain:J138) DNA.  
Chlamydia pneumoniae J138  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
1 (sites)  
Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,  
Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.  
Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CWL029 from USA  
Nucleic Acids Res. 28 (12), 2311-2314 (2000)  
2030349  
2 (bases 1 to 299650)  
Shirai,M.  
Direct Submission  
Submitted (04-Jul-2000) to the DDBJ/EMBL/GenBank databases.  
Mutsunori Shirai, Yamaguchi University School of Medicine,  
Department of Microbiology, 1-1-1 Minamikogushi, Ube, Yamaguchi  
755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp,  
Tel:81-836-22-2227, Fax:81-836-22-2415)  
On Aug 31, 2000 this sequence version replaced gi:6172286  
gi:6172288 gi:6172310 gi:6172312 gi:6172314 gi:6172316 gi:6172318  
gi:6172320 gi:6172322 gi:6172324 gi:6635158 gi:6635160 gi:6635162  
gi:6635164 gi:6635166 gi:6635168 gi:6635170 gi:6635172 gi:8547422  
gi:8547426 gi:8547430 gi:8978373.  
AB033780-AB033781, AB033792-AB033799: Submitted (25-Oct-1999)  
AB038345-AB038347: Submitted (14-Feb-2000)  
AB036071-AB036078: Submitted (18-Dec-2000).  
Location/Qualifiers  
1. 299650  
/organism="Chlamydia pneumoniae J138"  
/strain="J138"  
/db\_xref="taxon:138677"  
/note="synonym:Chlamydia pneumoniae (strain J138)"  
/gene="CpJ0001" .282)  
complement(1..282)  
/gene="CpJ0001"  
/codon\_start=1  
/transl\_table=11  
/product="CM001 hypothetical protein"  
/protein\_id="BA098211.1"  
/db\_xref="GI:8978374"  
/translation="MLGKIRGLSLVITLALNYGLIGITHNKLIITAKLCGVSTP  
ATQITVITIGIAGVLCILSPFCPSKSHSHGSDSCSGCHSHSDKN"  
573..878  
/gene="gacC"  
573..878  
/gene="gacC"  
/codon\_start=1  
/transl\_table=11  
/product="Glu-tRNA Gln amidotransferase (C subunit)"  
/protein\_id="BA098212.1"  
/db\_xref="GI:8978375"  
/translation="MEQFLDREETILLAKSAQLSEELIQEQTSLSAVTSKKEA  
LAIEDDADSCSESLFMHVVNVEDLREDSVTSDFNKEELIRNVPEISLGLVYPVAVIK"  
895..2373  
/gene="gata"  
895..2373  
/gene="gata"  
/codon\_start=1  
/transl\_table=11  
/product="Glu tRNA Gln amidotransferase (A subunit)"  
/protein\_id="BA098213.1"  
/db\_xref="GI:8978376"  
/translation="MYRSALSLAKAVTIGELTANGVQHPFHREEAGGVARISL  
CKEQLDQAEILDKRRSGEPGLKAGVPAGIKNDIHVTGLKTCASVLEIQPPFD  
ATVERIKEDGILLKLNDFAMGSTTLTSAFPIHPMDLSVPGSGGSAAY  
SARFCVPAIGDTGGSIRQAPAFGVGFKPSYGVASRYGVAFASIDQJGLPANTY  
EDVALMEFSGRDKDASREFFDSFMSKLTSEVPVIGVPRFELGLRDIEHNF  
FSSIALFEGEGHLVDLIDLSHVSITYILASAEATNLARDEGVGVGRSQPAHT  
ISQLYDSSRGSGFGKVMRRILLGNVYLSARQVYKKATAVRAKITYKARTAFERC  
EILAMPVCSAPAFELGTIDPVTLTLODTIVVAMALVLPALVAVPSGSKGELPLGLQ  
IIGGQGDQDQVGVGYSFQEHQAQIKLFSKRYAKSVVLGGOS"

gene  
CDS  
2370..3836  
/gene="gata"  
2370..3836  
/gene="gata"  
/codon\_start=1  
/transl\_table=11  
/product="(pet112) glu tRNA Gln amidotransferase (B  
subunit)"  
/protein\_id="BA098214.1"  
/db\_xref="GI:8978377"  
/translation="MSAYADWESYIGLEHVELNTASKLSSALNRGDEPNRIST  
VCTGLPGLPVLNOSAVKAVLFGCAVGEISLSLRFDRKSTFPDSPNRQVTFE  
PIIRGGRKAIYVGEERYEFLAOTIIEEDAGMLKHFEGAGVYNRAGVPIETKSP  
CMFCEPDAVAYATSLVSLDYIGISDCMNEGSIIRFDVNAVSRPSEPLNRKVEIK  
MNSFAPMADALEAKORQIDIEYNPNKDPVLVPAATYRNPDEKKIVLRKESAE  
DYKTRPEPDLPTLOQTESYIERIKTLEPDLKRYIKQYEGSLEDIASLISDKNI  
ATFEVFAKDKCKNFSLSNWSVEFGCKTLGVLPSSGTFPBGVAOLVNIIDQGI  
TKTAKELADCKNFSNPNPEILKEKPELLMSDEGLQKIIAEVLANPESIVDK  
NGKTKALFLVGOIKRTAGKAPKRVNELLILDKG"  
4127..6895  
/gene="pmp\_1"  
4127..6895  
/gene="pmp\_1"  
/codon\_start=1  
/transl\_table=11  
/product="polymorphic outer membrane protein G family"  
/protein\_id="BA098215.1"  
/db\_xref="GI:8978378"  
/translation="MRESLCGPPIVPSFTLLSVPTSLSATTTSLTPDSRFGDSOA  
ERSYVQAGDVSILTGDSVISNVDSNALNKKCFNVTSGSVTFAGNHGLVFNNTSSGT  
TRKGAVALCCQPOATVAFRSGFSTLFIQSPDIDKQGLYSKNAIMLNNVVRQEN  
QSKTKGALISGANVYVGNDSVSPYQNAATFGALHSGGLQILVNGAETVRFQNTA  
KNSGGALYSDDGDIIDQNAVYLPRENALITAIKGAVCCLPFGSGSTPVTFS  
DNKOLYPERNHSIMGGALIVARKLSSISGSGPLTPFNNTSVANSOMLGALINDGET  
SLSAKGTITTOGNRTSLPLNGTSHLONAKFLIQANGSIEPTDPTSDASTQ  
LNINDGPNKKEYTGITLFSGEKSLANDPRDFSTIPQVNNASAGLVITKEGAEVSK  
FTQSPGSHLVLDLGTKLJASKEDIAITGLAIDISLSSSTAAYIKATNAKQISVTD  
SIELISPTGNAYEDLRNRSQTFPLISLEPGAGGSVYTAGDPLVSPHFGQMKL  
AMTGKNGVGEFEFMKINYPKPEKEGALVPIILMGNAVDSLSMOVETHASLSQTD  
RGLMTDIGNFPHFVASEEDNIRYRHNSSGVLVYVNNETTPHRTAMASLOFELKAY  
AVSNNEVPMGYSTLYOYTTSGNIFRYSARNPANNVTLRRFLQNLMLFELICAY  
GHATNDKMTIDYANFPVAKNSMKNKNCMALECGSGMPLVFEKRLQOGAIPPKDLQVY  
AYQGFKEPTTADGRRFNSGLTSLISVPLGIREKALISQDVLVDFSFYIDIFKPD  
SCEALVYISGDSWILVPAHVSRAHFVGSGTGIRHNDYTELLRGSIECRHARVNYNI  
NCGSKFRP"  
complement(7138..7293)  
/gene="CpJ0006"  
complement(7138..7293)  
/gene="CpJ0006"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA098216.1"  
/db\_xref="GI:8978379"  
/translation="MQEPLRSALLERLSEMLVLLGVPSPETTRSTPEKANOLPKDSR  
NRTLESL"  
7515..10169  
/gene="CpJ0007"  
7515..10169  
/gene="CpJ0007"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA098217.2"  
/db\_xref="GI:9956083"  
/translation="MFSFLVVIPLTDDTSSLSLDEGNQSRKRIIAIYIAL  
SIIILAGGVNLTVAIPGLSVISSPAMGACALGCVMLAGIDVLAKKREVPVLAS  
VTTTPGVTSPPSSGISGADSTIRSLPTYLDEGHPQMRKIRIILAIYVLSIIIA  
SGVLLVVAIGLSIISSPAMGACALGCVMLAGIDVLAKKREVPVYVAPPIPEEY  
VIDIDDESILQOEAEALARPENMAFGYIKVVSLENNKMSLPYDGHGEKY  
KHQIRVRSLSKAWPEPDLIRIFEEFEFFLSAKRLIDLATIVYKRLITQOLER  
NNLRKAEVSYLVQDSIFKRIIDNEFLAKFMILSISCTPITIFENHEGVAKSLRH.  
NAVLEKVIYRSLOKSYRDIOMSSAKMILGNPFSLDENKKTIMKEHADMELSS

```

Db 5262 TGNAGGCTGATTTTCTGATATACACAGCAGAAATAGTGGAGTGGATTTATACCA 5203
Qy 1068 AAAGCTAGCCCTTCTCTGTGAGGCTTTACAGAG---TTTCTAGAAATATATGTCATC 1124
Db 5202 AAAATGCTGATTAATCTCAGACAGAGAGAGAGTATTTTCTAACAACAAGCTGGGAA 5143
Qy 1125 AGCAATCTTAAGGGGGTGTATCAGATGAGTGCCTCAGAGAGAGCTCTTCTGCG 1184
Db 5142 TGTCTACTCTTAAGAGAGGCAATTTGCAATTTCTGATTTCTGGAGAGATTTACATTTCTGC 5083
Qy 1185 AGACACAGAGAAACATTAACCTTTTGAAGAAATACCTTTACAACAACCGAGATACCGATAC 1244
Db 5082 AGATCTGGGCAATATCAATTTTTCGAGGCAATACACAGCCTACAGAGAGTCTCTGGAG 5023
Qy 1245 TCTTAACGTAATGCGATCAACATAGAGATACGAGAAATTCAGGAAATTCAGGAAATTCAGGCTGC 1304
Db 5022 TGTACACAGAAATGCTATAGATCTTCGATGCAATGCAAAATTTTAAATCTCCAGGAGAC 4963
Qy 1305 TAAAAATCATACATTTTCTCTATGATCCCATCCATCTCAGAAAGAACTCATCAGAGCT 1364
Db 4962 TCGGGGAAATTAAGTATTTTCTATGATCCTATCAGAGCTCAGGA--GCTACTGATAA 4906
Qy 1365 ATTGAATATTAATTAACGCTCTGCGGAGCTCTCATCATCATCAAGAAAGATTTCTAT 1424
Db 4905 GCTCTCTTTGATTAAGCTGACGAGAGATCTGAAATACCTATGAGGCTCATCTGTTT 4846
Qy 1425 TTTTGGAGAAACCTTAACAGCAGATGAATTAAGTTCTGACAAATTTAAATCTTCAT 1484
Db 4845 CTCTGGAGAGAACTCTCAGAG---TAAAGAACTGACAAATCTGAAGTCTACAT 4793
Qy 1485 CAGCAGCAGGCTCCCTATCCGAGAAAGTTATGCTCAAAAGAGAGACATTAGA 1544
Db 4792 TACACAGAGCTGTAGAGCTGTGCGAGGTGCCTTAGTATTTGAAGATGAGAGCTTAGT 4733
Qy 1545 GAGCAGAGGCTTCTCTCAAGAGGCGGTTCTCTCTGCGATCAGGATTCAGGAACGAT 1604
Db 4732 TGCAAATACATTAACGACGAGGTGAGGATGCAAGTGTATGATGAGAGGATCTACTT 4673
Qy 1605 ATCACTACAGCTGGGAGTATTTACATCAGAACTTAGAATCAATGTTGACTCTTAGG 1664
Db 4672 TGAAGCAAGGCTGAGGGGGTCACTCTCAATGCGCTAGCCATTAATATAGATTCCTAGA 4613
Qy 1665 TCTTAACGAGCCGCTGACCTAACGCAAAAGGTTCAAATTAAGTATGATCGTATCG 1724
Db 4612 TGGCAAAATTAACCTATTCATTAAGGCGACGCAAGATGAGATGTGCTTATACGG 4553
Qy 1725 GAAGCTCAACCTGATTAATTAAGAGGAACATTTATGAAGTCAATATGTTCAACCATGA 1784
Db 4552 GCGTATCATGCTGTATGATGCTCAGGGAACTATTAAGCATCATATATCTCACTCAACA 4493
Qy 1785 CCAAGCTCTCTCTATTAATAAATCACGGTATGATGATGTTGATCACTAAGTTCAT 1844
Db 4492 GCAGGTCTTCTTCTTAATAGCTTTCTGCAACAAGAACATGACTACAGATATCC 4433
Qy 1845 CAGAGGCTTATCCCTGTCTGTGAGATCCATATTCAGAAATACGATTCACAGACA 1904
Db 4432 CGATACCCCAATCTTAATCTACGATCACTAT-----GGGTATCAAGGGAAAC 4384
Qy 1905 ATGGAATGTAATTTGAGCTACGATACAGCTACAAATATCAAAAGAGCCAGCACTTG 1964
Db 4383 TGGAAATATTTGTTGGGTGACAGATGCACTGCAAAAACAATAAATCTACCTTAACCTTG 4324
Qy 1965 GACCAAAAACAGATTTGTTCCAGCCCGAAAGAAATCTGCGTTAGTATGCAATACCT 2024
Db 4323 GACTAATAACAGATACAAAGCGAATCCAGAACGTCAGAGACCTTTGGTCTTAATACCT 4264
Qy 2025 ATGGGAGGCTTACTACGAAATGCTCTGCAACACACTGTGAGATTCGGCGGAACCTGG 2084
Db 4263 GTGGGGTCTTTTGTGATGCTCGCTCATTCAGGCTTATGAGACCGGAGCAAAAGTTTC 4204
Qy 2085 TATGGAACACAAACAGGTTCTGGGTTCTCTCATGACGAACCTTCCGTACATAGACTGG 2144

```

```

Db 4203 GTTATCTTCTGCAACAATTTTGGGTATGAGATCGCGGACTTTTTCATGAAGATCA 4144
Qy 2145 AGATGAATATGCAAAAGGCTTCCGTATACCTGTGAGGCTACGTATCGGTGGAATGC 2204
Db 4143 GAAAGGAACCAACGATGATTAATCTCATTTAGCGCGGGTATCATTAAGAGAGATTT 4084
Qy 2205 TCACACTCTTAAGACACACTATTTACCTTTCCTTTCGCTATCTCTTTCTAGACAA 2264
Db 4083 CTTCAGGCTTCTGAATAATTTCTTAATTTTGTGTTTGTGACGTTTGTGGCTACGACA 4024
Qy 2265 AGATGTTTATGCTCACAACAACCTTAGAACCTACGAGTGAACCTTATTTCTTCAACA 2324
Db 4023 GGACCATCTTGTGCTTAAGAACCATACCATGATATAGCAGGGGCAATGAGATCCACA 3964
Qy 2325 CTCTCATACCTACAAACCCCAACTTTGAGATTAGAGAACCAAAAGTTTCTGATC 2384
Db 3963 C-----CTCGAGAGTCTAAGACCTCGCTAA 3937
Qy 2385 AGCTATGAATAATTTCCCTAGGAAATTTCCCTAGCCTTGGATGTCGAAGTTCTCAG 2444
Db 3936 GATTTTGTAGGAATTTCTGACTCCCTACCTTTTGTCTTCAATCTCGGTTGCTTATGG 3877
Qy 2445 CCATTACAGAACCGTATGGAACGCACTATACCTCATTTGCCAAGATCCGAAGGTTCTG 2504
Db 3876 CCATACCGACAATAACATGACCAAGATACACTGGCTATTTCTCTGTTAAGGAACTG 3817
Qy 2505 GAGCAACGAGTGTATACGTGTGTATCGGCTTAGACCTTCTTTTCTTCCAAACC 2564
Db 3816 GGGAAATGATGCTCTCGTATGAAATGTGAGAGAGCTATCCCGTATCTTCTCAGAGC 3757
Qy 2565 ACATCTCTTTTCAAGACCTTATTCACAGATGAAGTGAAGTGGTATATGATACAA 2624
Db 3756 TCGGCTTGTGGGTGATACCCACAGCCATTTCTAACCCTAGAGATGATATGACATCA 3687
Qy 2625 AAATAGCTTCTTCAAGCTCTAGTATGAGCGCTGTTTATTTGAGAAAGCTGCTTAA 2684
Db 3696 GAATGACTTTTAAGGAAACGACAGAGGCGCTTCTTCCAAAGTGAAGACCTCTTCAA 3637
Qy 2685 CCTCTCATCTCTGTGGGTGGAATTCGTGCAGGGGATATCGGAATTCCTACACTA 2744
Db 3636 CTTAGCGGCTCTGTAGGATTAATTT-----TGAATAATCTCCCATTAAGTCTACGTA 3583
Qy 2745 TGATCTCTCAGGATTTCTTGTTCGATGCTATGTAACAAATCCCAATCTACAGCGAC 2804
Db 3582 TGATCTCTCATACCTTACGTTCGCCATGTGATTCGTAATGATCCAGGCTGCACGACAAC 3523
Qy 2805 TCTTGTATGAGCCAGACTCTTGGAAATTTGCGGTGCGCAATCTTTCAGACAGGCAAT 2864
Db 3522 TCTTATGATTTCTGGGATTTCTGTGACATGHTACAAAGCTTGTCTAGACAACTCT 3463
Qy 2865 TTTTACTAGGGGTAGCAACAATACGTCTACACTCCAATTTGTAGCTTTTGGACATTTA 2924
Db 3462 TCTTGTACGTGTGGAATCATATGCTTGTCTTCAAACTTTGGAATTTTCACTGAGT 3403
Qy 2925 CGTATGAACTCCGTGATCTTCAAGAACTCAAAATGATGATGTTGTCACAACTCCG 2984
Db 3402 TGAAGTGAAGTGGAGGATCTTCTGTAGCTATGATGATCTTGGAGAGATTCGG 3343
Qy 2985 ATTCTA 2990
Db 3342 ATTCTA 3337

```

```

RESULT 15
AP002545 299650 bp DNA BCT 08-DEC-2000
LOCUS Chlamydomonas pneumoniae J138 genomic DNA, complete sequence,
DEFINITION section 1/4.
ACCESSION AP002545 AB033780 AB033781 AB033782 AB033793 AB033794 AB033795
AB033796 AB033797 AB033798 AB033799 AB036071 AB036072 AB036073
AB036074 AB036075 AB036076 AB036077 AB036078 AB038345 AB038346
AB038347 BA000008
VERSION AP002545.2 GI:9956082

```

	ORGANISM
REFERENCE	Chlamydia pneumoniae AR39 Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.
AUTHORS	1 (bases 1 to 12127) Read,T.D., Brunning,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., Deboy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M. Genome sequences of Chlamydia trachomatis Morn and Chlamydia pneumoniae AR39
TITLE	Nucleic Acids Res. 28 (6), 1397-1406 (2000)
JOURNAL	2 (bases 1 to 12127)
MEDLINE	10684935
PUBMED	
REFERENCE	Read,T.D., Brunning,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., Deboy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M. Direct Submission Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA On Jun 1, 2000 this sequence version replaced gi:7189672.
TITLE	Location/Organisms
JOURNAL	I. .12127
COMMENT	/organism="Chlamydia pneumoniae AR39"
FEATURES	/strain="AR39"
Source	/db_xref="taxon:115711" /note="synonym: Chlamydia pneumoniae AR39"
gene	complement(150. .2987) /gene="CP0757" /note-"This region contains an authentic point mutation, causing a premature stop, and is not the result of a sequencing artifact; similar to GP:4455886; identified by sequence similarity; putative:polymorphic membrane protein G family authentic point mutation" complement(3003. .3191) /gene="CP0758" complement(3003. .3191) /gene="CP0758" complement(3003. .3191) /gene="CP0758" /note-"hypothetical protein; identified by glimmer2; putative" /codon_start=1 /transl_table=11 /product-"hypothetical protein" /protein_id="AAE38560.1" /db_xref="GI:7189673" /translation-"MRLILSCPMFLIACASFEFGDEMOGRINQSIDNASSIGELF SISTKGYSCLELHREIR" complement(3336. .6086) /gene="CP0759" /note-"This region contains a gene with one or more premature stops or frameshifts, and is not the result of a sequencing artifact; similar to GP:4455886; identified by sequence similarity; putative:polymorphic membrane protein G family, degenerate" complement(6325. .9110) /gene="CP0760" /note-"This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GP:4455886; identified by sequence similarity; putative:polymorphic membrane protein G family, authentic frameshift" complement(9393. .11918) /gene="CP0761" complement(9393. .11918) /gene="CP0761" /note-"similar to GP:4376270; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product-"polymorphic membrane protein G family" /protein_id="AAE38561.1"
gene	
CDS	
gene	
CDS	

		/db.xref="GI:1189674"	
		translation="MKRLPLRLISLIVPLRLMSNLGLAATPEELSLNSPDCTSTPS	
		FSSRTSADIDETNVERKDSVYIEHPVPTGTGETOSTSCRKNDPAADLFLCGGSRFTPS	
		NIDMTASGAIAIGSEAAKNTVLLSGFSAFLSFLSPASTVINGALIAVKNLSLIDLD	
		KVLIDDNFSTGDGCAINCAGSLKLIANKSLSPFIONSSRTGALAHNTNLSISGELT	
		FOGRTAPPAAGKGAIAIADSGTSLISGDSGDILFEENGTGATVSHSLIDGTSK	
		ITLARAOGHTIYEPDITLVYGTSTVADALININPDIDGNKEYGCTLVFGEKLTAEI	
		AKDKNRTSKLLOWAFKNGTVYKLGQVIVLSANFESODANSKILMDGTSLIVANESI	
		ILEPMLNIDISIRNGKRIKLSAATAOKDIDILDRPVALTIDESQONGFLPNEHSYDG	
		IILEDKARDIVISDSDSIDAVSPRYGCKRTINMSTDOKKATVSNAKOSPNPRAQ	
		EAPLVPLMLWSFIDVRSFONFIELEGEARKEPFWNVAJISVLSLRLEHDSLSVY	
		VSGAVVVGASTRMFGDPTLSLGEAQLAROKDFYMNRLVAGTSLRLDHDLSLVY	
		GYSIILGSGGRLREILLPYVRSKTLPCSFPGOLSYGIVADIRRMKTESLAPPLSLDHTSM	
		GVYIAGAGELTRERAVENTSGRSGFPOEYGTPEVKQAVYARDBOFEVLALSRDSESHL	
		YNLALPIGLIKERAEADYYHYVAMVSPDORSNPKCTTLLISNOGSMKTKGSLNARO	
		AGIVASGFRSLGAATLFGNFGEMGSSHSYVNDGSKIKP"	
		1 others	
BASE COUNT	3532 a	2722 c	2431 g
ORIGIN			3441 t
Query Match	15.0%;	Score 479.2;	DB 1; Length 12127;
Best Local Similarity	51.4%;	Pred. No. 1.2e-90;	
Matches 1390;	Conservative 1;	Mismatches 1234;	Indels 81; Gaps 9;
OY	297	TGATGATAGCTTTAAATGGAATATGATTCAGGAACGTTTACTCCAAAACCTTCAGCCAC	356
Db	5973	TGCTATATCAAGACACACAGATTTTACTCTAAACGGGCACACTTCGTAGTCATGAGCAC	5914
OY	357	AACATATCTCTAAAGAGATGCTCTCTTTTACGAGCTGTGAAAAGGACATCCCTCTATC	416
Db	5913	GACCTATATTCCTCGATGGGAGTGTCTGATTAAGCCACAGAGGAAACAAACAGACTTAA	5854
OY	417	TGACAGTGTGTTTAAAGCAAAACCCAGGACATCTTACCTTCTTGGGAAACGGTCATAGCTT	476
Db	5853	CACAGATGCTTTTCTTACACATCGCAGAAATCTTACCTTTTAAAGGAACGATTTTCTCT	5794
OY	477	AACGTTTGGCTTTATAGATGCTGCGACATCAGCAGGTGCTGCTG---CATCTACAAACG	533
Db	5793	TCATTTTACAAATATATTATTTGGCTACAGGTGGCAGAGTGTGTGTTAGCAATACGAGCAG	5734
OY	534	AAATAGATATCTTACCTTCTCAGGGTGTTCCTTCTACTGAGTTTGATTTGATTCCTCTCAGAC	593
Db	5733	TTTCGGGATTAAGAAATCTCAGAGATTTTCAACCTTTCGGATGCTTGCAGCTCTAG---	5675
OY	594	AACGGTTCACAGGTCAAGGACGCTTTCCTCAGCAGAGAGCGTAAATTTAGAAATAT	653
Db	5676	-----GACCAACAGGTAAAGAGCGCATTAATTAACGATGCTGCTGCTGTGAGACGAT	5623
OY	654	TGCTAACTCTGAGTTCGTGGGAATTTTCTACTGCAGATGCTGAGAGCTATCAAAAGAGC	713
Db	5622	AGGGAATCTTGATCTTAATGAATATGCTCTATGTAATAATGGGAGGACCATTAATACAA	5563
OY	714	GTCCTTCTCTTAACTGGACCTCTGAGATGCTCTTTTATGTAACAATCTTCATCAAC	773
Db	5562	GACTTGTCTTGAATCGGAGACGCGGTTGTATGACGTTTCTTGGCATACTCTCTGCCA	5503
OY	774	AAAGGAGAGACAATGCTACTACAGCAGGCGCTTCGATACCAATATACAGGTTATGT	833
Db	5502	ACAAAGGGGAGATCTATGCTCTGAGCTCTGTGATTTCTTGAAATVGAAGGAATCTT	5443
OY	834	TAGATTCCTATCTAAATAGCGGTCTACGTCAAGAGGCGCTATCGATGTGAAGCACATC	893
Db	5442	GAGCTTGGGAACAACAATGCTGGACAACATCAGGAAGGCGCATCTCTGTAAGGGAACCT	5383
OY	894	GATCTATTCGAACAACAATTTCTATATTTTGAAGGGAATGCAGGAAACTACTGCGGG	953
Db	5382	TGTGATCTCCATAACCAAAATATCTTTTGTGATGGCTGCAAAAGCACTTAAATAATGGGG	5323
OY	954	TGCGAT-----CTGCAACACCAAGGCGAGTGGATCTCTCAACGATTAATCTTAAACA	1007
Db	5322	AGCTTATTTATGTGAACAAGCAGAGGGGCAACCCAGACCTCTATCTTGACTCTTTCAAGGAA	5263
OY	1008	TAAAGCTCTGATCTTTGCTTCAAAACGTAGCAAAACAGCGGTGGCGCCATCTAGCTAA	1067

Oy	1020	CTTGTCTTCAACAGCTACAGAAACAAAGCGGTGGCGCATCACTCAAGTAAAGACTAGCCCT	1079
Db	2930	TTTTGGAGAAAATACTCTTCTTCAGAAAGGTGGAGGATTTACACCGATTAATCATATAT	2989
Oy	1080	TTTCTCTGAGAGCTTATACAGATTTTCTACGAATTAATGTCTATACGAACTCTAAAGG	1139
Db	2990	GACTTCTGGTGGCGCTTAGGCCATTTATCAATTAACAAAGTTACCGATGCTACACCTTAAGG	3049
Oy	1140	GGGTGCTACGACATCGATGCTCTCAGAGAGCTCAGTCTTTCTGCAGACAGAAACAT	1199
Db	3050	TGGAGCTATTTGGTTGTGCTGCAATGAGAGAAATGTACTTTAACCGCTGAACATGGGGATAT	3109
Oy	1200	TACCTTTCTAAGAAATACCTCTTAACAAACCGGAABTACCCTACTCTCTAAACGTAATGC	1259
Db	3110	TACTTTTG--ATATATTAACCTGTATGGCCACACAAGCAATGGTACATATAAAAAAGAAATGC	3166
Oy	1260	GATCAACATATAGAAAGTAACGGAAATTCACGGAATTAACGGGTGCTAAAAATCATATACAT	1319
Db	3167	CATTATACATTGAAGGCAATGGTAAATTTCCGTCAACTTACGTGCAGCGTCTGGAAGACAT	3226
Oy	1320	TTTCTTTATGATGCCATCACTTCAGAAAGAACCTTCAGACGTAATGAAGTAAATTA	1379
Db	3227	TTCTTTCTATGATCTATCACTACAGTTGAAGTAAATGCTGTGATCTTCACCTTTGAATAA	3286
Oy	1380	CGGCTCTCGGGAGCTCTCAATCCATATCAAGAACGATTTCTATTTTCTGGAGAAACCT	1439
Db	3287	-----AGCTGAGGGTATATAAACGTAATATGMAAATTAATTTTTCAGGAGAAACCT	3340
Oy	1440	AACAGCATATCACTTAAATCTGCTGCACATTTAAATTTCTATTACAGCACCCAGTCTC	1499
Db	3341	CACGTGMAAGAACACCTCTGTGTGGGATTAACCTTAAGAACAAATTTACACACCGCTATAC	3400
Oy	1500	CTTATCCGAGAGAAAGTTATTTGCTACAAAAGGAGTCACTTTAGAGACACGAGCTTCTC	1559
Db	3401	TTTATGCTGCTGTGTAACCTTTGTTTACGAGCGGTGGAGATAGAAACAAAACAGTGGT	3460
Oy	1560	TCAAGAGGCGGTTCTCTCTCGGCATGATTTCAGAACAGCATTTATCACTACAGCTGG	1619
Db	3461	GCAAAACAGCAGACTTTTGATTTCTGTATGATGAGCGCACAAAGTTATCCGCAAAAACGA	3520
Oy	1620	GAGTATTTACATCAACGAACCTTAGAATTCATGTGTACCTTAGGCTTAAAGCAGCCCGT	1679
Db	3521	AGATGCTACACTGACGAATGGCTATTATATCCGAATACCTTAGATGGGAAAAAATTGC	3580
Oy	1680	CAGCCTAACAGCAAAAGTGTCTTCAATTAAGTATGATGTAATGGGAAGCTCAACCTGAT	1739
Db	3581	CGTAGTGCATGCCGTCTGCTCGGAAAGATGATGACTTATATCAGGTGCTATTTGGGGTAT	3640
Oy	1740	TGATATTTGAAGGAACTTTATGAAGTATATGTTGACCCATGACACAGCTTCTCTCT	1799
Db	3641	TGATCTCACAGGGGAAGTTTTATGAAAAACATTAAGCTAAATGATATAGTTAGGAGG	3700
Oy	1800	ATTAAAAATACGGTGTGATCTGATGTTATCTAATCACTGTATACATCAGACAGCCTTATCC	1859
Db	3701	AATTCACACTTTCTGGAAAGGTGCGGTGACMC-----AACCAACGTGCTAG	3748
Oy	1860	TGTTCTCTGAGAGATCTTATATTCAGATATGCGATTCCAGGACAAATGGAAATGTAATG	1919
Db	3749	TCATGTTGTTGGTGTGTGCTTAACCCACTATGTTATATCAAGAAACTGCTGTACAGTTG	3808
Oy	1920	GACTACGGATA-----CAGCTCAAAATACAAAAGAGCGCACGCAACTTGGACCACAAAC	1973
Db	3809	GGTCAAAAGATTAATACCTGTATCTTAAAAACAAACAGCAAACTTTTACTCGAATTAACAC	3868
Oy	1974	AGATTTGTTCACAGCCCCGAAGAAATCTCGGTATGATATCACTTACCTATGAGGGAGT	2033
Db	3869	AGGATATGTTTCCAAATCTGAAAGCTGTGCTCCGCTAGTACTCAATATACCTTTGGGGATC	3928
Oy	2034	CTTATCTGACATTTGCTCTCTGCAACAGCTTTAGAGATCGGCGCAACTGTGATGAAACA	2093
Db	3929	CTTATATGATTTACGTTCTTATTCAAATATGTCCTTGGAGCTAGTGTGATATGATTTCTTGA	3988
Oy	2094	CAAAACA---GGTTTCTGGGTTCTTCATGACGAACCTTCTGCATTAAGACTGAGATGA	2150

Db	3969	GACAGCTGCTGGTTTGTGGGCTCTCGAATTTGGGAACCTTCTCCATGAAGATCGGAATGC	4048
Qy	2151	AAATGCAAAAGGCTTCGTCATACCTCTGGAGGCTACGTATCGGTGAAGTGTCAAC	2210
Db	4049	TGAAATATGCAAAATTCGTCATATCACTTCGGGATATGTGTTAGAGCCACAAACAAATAC	4108
Qy	2211	TTCCTAAGACACCTATTATTACCTTTGGTTCCTCCACTCTCTTTGGCTAAGACAAAGATTG	2270
Db	4109	CTCGAGAGGATTCCTCTGTAGTGGCTTCTCTCGATTATTTGCAAAAAGATTAAGACTA	4168
Qy	2271	TTTTATGCTTCAACAACACTAGAACCTACGCTGGTGAACCTTATATCTTCAACACTCTCA	2330
Db	4169	CCTTGTAAAGCAAGAACGCCCAACCGCTATGGGGTCTCTGATATATTACACAGATGTAG	4228
Qy	2331	TACCTTACACCCCAAAACTATTGTGAGATTAGAGAGCAAAAGTTTCTGAATCAGTAT	2390
Db	4229	CAAGTTTATGATCTCACGGGGTATTATTAATGGGCTTACACAGTGTGTTCAAGG----	4283
Qy	2391	AGAAAAATTCCTTGGGAATTTCCCTAGCCTTGATATGTCACAGTTTGGTTCAGCATTC	2450
Db	4284	-----TTTCTTAAAGATTTCTTATTTCTTATTTCTTGAGTACAAATTAATCTATTGGCACAC	4336
Qy	2451	AGACACAGTATGGAAGACGACTATACCTCATTCGACAGATCCGAAGTTCCTTGAGCA	2510
Db	4337	GGCCAACACATATGACACAGTCCTATACAGCTATCTGTGAAGTGAAGAGTTCCTTGGGATA	4396
Qy	2511	CGAGTATATAGCTGTGGTATCGGCTACAGCTTCTTTGTTCTTCCAACCCACATTC	2570
Db	4397	TGATATACCTGGGCTTAATCTTGTCTATAGCGTACCTATCCCGGATTTAGTTC---TTC	4453
Qy	2571	TCTTTTCAAGCCCTTCATTCACAGATGAAGTCGAATAGTATTATGTCACAAATAG	2630
Db	4454	TATCTTGTATGTTATATCACCCGTTTGCAAAATTACAAGTTGTCTATGGCACCAAGATGA	4513
Qy	2631	CTTCTCGAAAGCTCTAGTATGAGCGGTGTTTACTATTGGAAGGCTGCTTAACCTTC	2690
Db	4514	CTTTAAAGAACCAACACAGAGAGCGGGCTTTGAAACACACGATCTTCTCAACGTTTC	4573
Qy	2691	GATTTCGTGGGTCCGAATTCGTGCAGGGGGATATCGAGATTCCTTACACTATGATCT	2750
Db	4574	TGTACCTATAGGATTAATAAATTTAGAAACCTCTCCTATGAGAGAGAAAGTGTATTAGACT	4633
Qy	2751	CTCAGAGATCTTTTGTCCGATGCTATGCTATACATTCGCCAATCTACAGCACTCTGT	2810
Db	4634	TACACTGATGATATACCTGATGTGTACCGCTATTAATCCAACTGTATAGACAGATTGGC	4693
Qy	2811	GATGAGCCACAGACTCTTGAAATTTGCGGGTGGCAATCTTTCAGACAGCAATTTTACT	2870
Db	4694	GATCAATGACGTTTCCGTGTTAAACACAGCTACGATGTGCTTAACAAAGCTTTCATAGT	4753
Qy	2871	GAGGGTATGACAAACATAGCTTCAACATCCCAATGTGTAGGCTCTCGGCACTTAACGAT	2930
Db	4754	TGCGGCGGGTACCAATTTCCCTTAACCTGTGGTGTGAGATGTTCAAGTATTTGGTTT	4813
Qy	2931	GGAATCCGTGATCTTCAAGAGACTACATAGATGTGTGGTACCAAACTCCGATCTTA	2990
Db	4814	CGAATTAAGAAAGCTCTTCAGAAATATATAACGTAGATCTTGGCGCTAAGTGTGCGTCTA	4873
Qy	2991	GATTGCT 2997	
Db	4874	AAAAGCT 4880	

RESULT	14	
AE002235/c		
LOCUS	AE002235	12127 bp DNA
DEFINITION	Chlamydia pneumoniae AR39, section 63 of 94 of the complete genome.	
ACCESSION	AE002235	AE002161
VERSION	AE002235.2	GI:8163495
KEYWORDS	Chlamydia pneumoniae AR39.	
SOURCE		





QY	834	TAGATTCTATCTAACATAGCGTGTACGTAGAGAGCGCTATCGATGATGAAGCAGCTC	893
Db	645	GAGGTCGGAAACAACAGTGGACACATCAGAGAGCGGATCTCTGCGAAGGGAACCT	704
QY	894	GATACTATTCGAACAACAATTTCTATTTTGAAGGGAATGCGGCAAAATACAGCGGG	953
Db	705	TGTGATCTCCAAATTAACCAAAATATCTTTTGCATGGCTCGAAAGCACTCAAAATGGCGG	764
QY	954	TGCGAT-----CTGCAACACCAAGGCGATGATGCTCTGAATGTATCTCAACAA	1007
Db	765	AGCTATTGATTGTAAACAAGAGGGGGGAACCCAGAACCTATCTTGACGCTTTGAGGAAA	824
QY	1008	TAGACGTCGTATCTTTGCTTCCTCAACAGTACGAGAAACAGCGGTGGCCCATCATGCTPA	1067
Db	825	TGAGAGCCTCGATTTTCTGATTAACACAGCAGCAAAATAGTGGAGGTGCGATTTATACCAA	884
QY	1068	AAAGCTACCCCTTTCCTCGTAGAGGCTTTACAGGTTTACAGAAATTAATGCTCATC---	1124
Db	885	AAATTTGGTGTATTATCTCAGAGAGAGAGAGAGGATTTATTTCTCAACAAAAAGCTGCANA	944
QY	1125	AGCAACTCTAAGGGGGGTCGTATACAGCATCAGTGGCTCAGAGAGAGCTAGTCTTCGC	1184
Db	945	TGCTACTCTTAAGAGAGGGGCAATTTGGGATCTAGATTTGGAGAGATTAGCATTTTCGC	1004
QY	1185	AGAGACAGGAACATTTACCTTTGTAGAATACCCTTACACACCGGAGTAGCCGATAC	1244
Db	1005	AGATCTCGGCAATATCATTTTTCGAGGCAATACAGACACTACAGGAAGTCTCGGAG	1064
QY	1245	TCTCAAGCTAATGCGATTCACAATAGCAATACGGGAATTCAGGGAATTCGGGGTGC	1304
Db	1065	TGTGACCAAGAAATCTTAGATCTTGATCGATGGAATGCAAAATTTTAAATCTCGAGGAC	1124
QY	1305	TAAAAATCATCAATTTTCTTATATGATCCCATCACTTGAGAGGAACCTCATAGAGCT	1364
Db	1125	TGCGGGAATATAAGTATATTTTCTATGATCTCATACAGAGCTCAGGA--GCTACTGATPA	1181
QY	1365	ATTGAAGTAAATTAACGGCTTCGCGGAGCTCAATTCATATCAAGAGACATTTCTTT	1424
Db	1182	GCTCTTTTGAATTAAGCTGACCGCAGGATCTGGAATTAACCTATGAAGGCTACATGTTTT	1241
QY	1425	TTTCTGAGAAACCTTAACAGAGATGAGCTTAAGTTGCTGACAAATTTAAATCTTCAT	1484
Db	1242	CTCTGGAGAAACTCTCAGAGAGGAACCTTAAGAACTGCAACATCTGAATCTTCAAT	1301
QY	1485	CACGAGCAGCTCTCCCTATCCGAGGAATTTATGCTACAAAAGGAGCTACTTTAGA	1544
Db	1302	TACACAGGCTGTAAAGTGTCTGAGGTGGCTTAGATTAAGTAAGATGAGATGACTGTAGT	1361
QY	1545	GAGCAGGAGCTTCTCTCAGAGGCGGTTCTCTCTCGSCATGATTCAGAAAGACATT	1604
Db	1362	TGCAAAATACATTAACGAGGTGCAAGGATGGAAGTCTTTAAGATGAGGAGGACACTTT	1421
QY	1605	ATCAACTACAGCTGGGAGTATTTACATACAGAACCTAGAGATCAATGTTACTOCTTAG	1664
Db	1422	TGAGGCAAGGCTGAGGGGCTCATCTCAATGCGCTAGCGCATTAATATAGATTTCTTGA	1481
QY	1665	TCTTAAACAGCCCTGCAGCTTAACAGCAAAAGGTGCTCAATTAAGTGATGATTTGCG	1724
Db	1482	TGGGACAATTAAGCTATCATTTAAGCGAGAGGAGCAATTAAGATGTTGCTTATCAGG	1541
QY	1725	GAAAGCTCAACCTGATGTATTTGAAGGAAACATTTATGAAGTCAATATGTTACAGCATGA	1784
Db	1602	GCAGGCTTTCTTTAATTAAGCTTCTCACAAGGAACGATAGCTACTACAGATATCC	1661
QY	1845	CAGAGCCCTTATCCCTTCTCTGCTGAGAGTCCCTAATTCAGAAATTCGGATTCAGAGCA	1904
Db	1662	CGATACCCCAAT-----TCTAAATACAGGAATCTACTTGGGATATCAAGGAGAC	1709
QY	1905	ATGGAATGTTAATGAGTACGGATACAGTACAGTACAAATACAAAGAGGGCACGGCAACTTG	1964

Db	1710	TGGAAATAATCTTTGGGTCGACGATGCAACTGCGAAAACAAAAAGCTACCTTAATCTTG	1759
Qy	1965	GACAAACACAGAGATTGTGTCACACCCCGGAAAATAATCTGCGTATGATGCATACCCCT	2024
Db	1770	GACTAAACACAGATACAAAGCGAATCCAGAAAGTCAGAGACCTTTGGTTCTCTAATAGCCT	1829
Qy	2025	ATGGGAGCTCTTACTGACATTTGCGTCTGTGCACACGCTGTAGAGATCGCGCACTGG	2084
Db	1830	GTTGGCTCTTTTGTGTCGATGCCCTCATTCAGACCTCTCATGACCGGAGCACAGTTTC	1889
Qy	2085	TATGGAACACAAACAGTTTCTGGGGTTCCCTCATAGCAAGACTCCGCAATAACAGCG	2144
Db	1890	GTTATCTTGTCACAAATTTGGGGATCAGAAATCGGGACTTTTTCATGAAGATCA	1949
Qy	2145	AGATAAATATGCAAAAGGCTTCGCTCATCTTACCTTGGAGGCTACGTCATGCGTGAAGTGC	2204
Db	1950	GAAAGGAACCAACAGTAGTATATGCTCATTTCTACGGGGGTATGCAITTAGAGAGAGATT	2009
Qy	2205	TCACACTCTAAAGACGACTATTTACCTTTGCGTTGCTGCCATCTCTTTGCTAGACAA	2264
Db	2010	CTTACAGGCTCTCGAAAATTTCTTAATTTTGGTTTGTGACGCTTTTGGCTACGACAA	2069
Qy	2265	AGATTGTTTATCGCTACACAACTCTAGAACCTTACAGCTGAGTGAATTTATTTCTTAACA	2324
Db	2070	GGACCATCTTGTGGCTAAGAACCTTACCACATGATATAGAGGGGCAATGATTTACCACA	2129
Qy	2325	CTCTCATACCTTACACACCCCAAACTATTTGAGATTAGAGAGCAAGTATTTGCAATC	2384
Db	2130	C-----CTCGAGAGTCTAAGACCCCTGCTAA	2186
Qy	2385	AGCTATAGAAAATTCCTTACGGAATTCCTCCATAGCCTTGATGTCACATTTGCTTACG	2444
Db	2157	GATTTTGTACAGAAATTTCTGACTCCCTACCTTTTGTCTTCAATGCTCGGTTGCTTATGG	2216
Qy	2445	CCATTCACACAAACGGTATGGAAGACGCTTACTTATTTGCCAAGATCCGAAGTTCTTG	2504
Db	2217	CCATTCGACAAATACATATGACCAACAAAGTACATGCTATTTCTCCTTTAAGGGAAGCTG	2276
Qy	2505	GAGCAACAGATGTACTCGGTGGTATCGGCGCTAGACCTTCCTTTGTCTTTCCAAACC	2564
Db	2277	GGGAAATATATGCTTCGGTATAGATATGAGAGAGCTATCCCGTATGTTCTTACGAGCG	2336
Qy	2565	ACATCCTCTTTTCAAGACCTTCATTCACAGATGAAGTGAATGGTTATGTATATACCA	2624
Db	2337	TGCGCTTGGGTGATACCCACAGGCCATTTCTAAACCTAGAGATGATCATGACACATCA	2396
Qy	2625	AAATAGCTTCTTCAAAAGCTCTAGTATGGCCGTGTTTATGATTTGGAAGCTGCTTAA	2684
Db	2397	GAAATACCTTTAAGAAAAACGGCACAGAGGCCCTTCTTCCAAAGTGAAGACCTTTCAA	2456
Qy	2685	CCCTCGATTCCTGCTGGGTGGGAAATTCGTCAGGGGGATATCGGAGTATCCTCAACCTA	2744
Db	2457	TCTTACGGTTCCTTATAGGATAAATTT-----TGAAATTTCTCCGATTAAGTCTACGTA	2510
Qy	2745	TGATCTCAAGAGATTCTTGTTCGATGCTATTCGTAACAATCCCAATCTACAGCGAC	2804
Db	2511	TGATCTCTCACTTACCTTACGTTCCGATGATGTATGTATGATTCACAGCGCTCAGACAAAC	2570
Qy	2805	TCTTGTGATGAGCCACACCTTGTGAAAATTCGCGGTGGCAATCTTTCAAGACAGCAATT	2864
Db	2571	TCTTATGCTTCTTGGGATTTCTTGGTCGACATGCTGTACACCTTGTCTAGACAAAGCTCT	2630
Qy	2865	TTTACTGAGGGGTGCAACAACTATGCTTACAACTTCATTTGAGGCTCTGCGACATTA	2924
Db	2631	TCTTGTACGTCCTGGAATCATCATAGCTTGTGCTTAAACCTTTGAAGTTTCTACGCACTT	2690
Qy	2925	CGCTATGGAACCTCGGTGATCTTCAAGAGACTACATATGATATGTTGGTACCAACTCCG	2984
Db	2691	TGAAGTCAGATTTGGAGAGTTCTTCTGTAGACTATGCTATGCATCTTTGGAGGAAGTTCCG	2750
Qy	2985	ATTCCTA 2990	

```
Db 22387 GCAGGCTCTTCCCTTAATAGAGCTTCTGCACAGGAAGCATGACTACTACATATATCCC 22446
Oy 1845 GAGCAGCCTTATCCCTGTTCCGCTGAGAGTCTTAATAGAAATACGATTCACAGGACA 1904
Db 22447 GCATACCCCAATTCCTAAATACGATACATCTAT-----GGGATCAAGGGAAAC 22495
Oy 1905 ATGGAATGTTAATGAGTACGATACAGCTCAATAATACAAAAGAGCCAGCGCACTTG 1964
Db 22496 TGGAAATTAATGTTGGTGCGAGATGCAACTGCAAAACAAAATGCTACTTAACTTG 22555
Oy 1965 GACCAAAACAGAGATTGTTCCAGCCCGGAAAAGAAATCTGCGTATGATGCAATACCT 2024
Db 22556 GACTAAACAGAGATACAAAGCCGATTCAGAAAGCTGAGGACCTTGGTTCTTAATAGCCT 22615
Oy 2025 ATGGGAGCTTACTGACATTCGCTCTCTGACACAGCTTGATGAGATGGCGCACTG 2084
Db 22616 GTGGGGTCTTGTGTCGATGCTCCATTCAGAGCTCAGAGCCGAGGACCAAGTTG 22675
Oy 2085 TATGNAACAACAAGGTTCTGGGTTTCTCATGACGAACTTCGCTAATAAGCTG 2144
Db 22676 GTTATCTTCGACAAATTTGTGGGTATCAGGAATCCGGGACTTTTGCAATGAAGATCA 22735
Oy 2145 AATATGAAAATCCAAAGCTTCGCTCATCTCTGAGGCTACGTCATCGTGGAATGTC 2204
Db 22736 GAAGGAAAACCAACGTATGATCTAGCGCGGCTATGATGATGAGAGAGGAT 22795
Oy 2205 TCACATCCTTAAGAGAGACCTTATACCTTTCGTTTCCATCTCTTCTGACAGACA 2264
Db 22796 CTTACAGGCTTCGAAAATTTCTTTAATTTTCTTTTGTCTGAGCTTTTGGCTACGACA 22855
Oy 2265 AGATGTTTATCGCTCACACACACTAGAACCTACGGTGAGACTTAAATCTCAAGCA 2324
Db 22856 GGACCATCTTGTGGCTAGAAACCATACCATGATATGAGAGGGCAATGATACGACA 22915
Oy 2325 CTCTCATACCTACCAACCCCAAACTATTTGAGATTAGGAAGCAAGATTTCGATTC 2384
Db 22916 C-----CTCGAGAGTCTAGAACCCCTGCTAA 22942
Oy 2385 ACCTATAGAAAATTCCTCAGGAAATTCCTCAGCTTGGATGTCGAAATTCGTTGAG 2444
Db 22943 GATTTTGTGAGAAATTCGACTCCCTTCTTGTCTTCAATGCTCGGTTTGTCTTAAG 23002
Oy 2445 CCATTCAGAACCTATGAGAAACGCACTATACCTATTCGCAAGATCCGAAGTTCTTG 2504
Db 23003 CCATTCAGAACCTATGAGAAACGCACTATTCCTGTTAAGGAAAGCTG 23062
Oy 2505 GAGCAACGAGTATAGCTGAGGAGTATGAGCTTCTCTTCTTCTTCTTCAACCC 2564
Db 23063 GGGAAATGATGCTTCTGATGAAATGAGAGACTATCCGCTGATGCTTCAAGAG 23122
Oy 2565 ACATCCTCTTCAAGACCTTCAATTCACAGATGAAAGTGAATGGTTATGATACCA 2624
Db 23123 TGGGCTTGGGAGTATGACCAACGCACTTTTAAACCTAGAGATGATCTATGCAATCA 23182
Oy 2625 AATATGCTTCTTGAAGCTTATGATGAGCCGTTTATGATTTGAAAGGCTCTTAA 2684
Db 23183 GAATGACTTTAAGAAAAGGACAGAAAGCGCTCTTCCAAAGTGAAGACTCTTCAA 23242
Oy 2685 CCTTCGATTCCTGAGGAGGAAATTCGTCGAGGGGATATCGAGATTCCTAACCTA 2744
Db 23243 TCTTACGGGTTCTCTAGGATGAAATTT-----TGAGAAATTCCTCCGATTAAGTCAAGTA 23296
Oy 2745 TGAATCTCAGAGATTTCTTGTTCGATGTCATGTAACATCCCAATCTACAGGAC 2804
Db 23297 TGAATCTCAGAGATTTCTTGTTCGATGTCATGTAACATCCCAATCTACAGGAC 23356
Oy 2805 TCTTGTATGAGCCAGACTTCTGAAAATTCGCGGTGCAATCTTTCAAGACGCAAT 2864
Db 23357 TCTTGTATGAGCCAGACTTCTTGTGTCGACATGTCGTAACGCTTGTCAAGACGCTCT 23416
Oy 2865 TTTACTGAGGGGTACCAACAATAGCTACAACTCAAACTGTGAGCTCTTCGACATTA 2924
Db 23417 TCTTGTAGCTGTGGAATATCATGCTTGTGCTTCAAACTTTGAAGTTTTCAGTCACTT 23476
```

```
Oy 2925 CGCTATGGAACCTCGGTGATCTTCAAGGATACATGATGATGTTGGTACCAACTCCG 2984
Db 23477 TGAAGTCGAGTTGCGAGGTTCTTCTGTAGCTATGCTATGATCTTGGAGGAATTCGG 23536
Oy 2985 ATTCCTA 2990
Db 23537 ATTTTA 23542

RESULT 12
A81837
LOCUS A81837 2757 bp DNA
DEFINITION Sequence 11 from patent WO9858953. PAT 21-JAN-2000
ACCESSION A81837
VERSION A81837.1 GI:6731869
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 2757)
AUTHORS Madsen,A. and Birkelund,S.
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL Patent: WO 9858953-A 11 30-DEC-1998;
MADSEN ANNA SØFIE (DK); BIRKELUND SVEND (DK)
FEATURES
source location/Qualifiers
1..2757
/organism="unidentified"
/db_xref="taxon:3264"
BASE COUNT 770 a 564 c 642 g 781 t
ORIGIN

Query Match 15.6%; Score 498.4; DB: Length 2757;
Best Local Similarity 51.6%; Pred. No. 1.1e-94;
Matches 1395; Conservative 0; Mismatches 1236; Indels 75; Gaps 8;

Oy 297 TGATGATAGCTTTAATGGAATATGATTCAGAGGAGTTTACTCCAAAACCTTCAGCCAC 356
Db 114 TGTGATATCAAGACACACAGATTTACTTCCTTAAGCGGCAACTTGTGATGAGGAC 173
Oy 357 AACATATTCCTTACAGAGAGATGCTCTTTTACGAGCCTGGAAGGACACTCCCTATC 416
Db 174 GACCTATATTCGATGGGAGTCTCGATTAAGCCAAAGGAAACAAAGAGCTTAAC 233
Oy 417 TGACGATGTTTAAAGCAACGAGGACAAATCTTCTTGGGGAACGTCATAGCTT 476
Db 234 CACAAGTTGTTTCTTCAACACTGACAGGAATCTTACCTTGAAGGAAACGATTTTCT 293
Oy 477 AACGTTGCTTATAGATGCTGGACATCATGACAGTCTGCTG---CATCTACAAGAC 533
Db 294 TCATTTGACATATATTTTCTCTACGTCGTTCAGAGTGTGTTGTTAGCAATACAGCAGC 353
Oy 417 TGACGATGTTTAAAGCAACGAGGACAAATCTTCTTGGGGAACGTCATAGCTT 476
Db 234 CACAAGTTGTTTCTTCAACACTGACAGGAATCTTACCTTGAAGGAAACGATTTTCT 293
Oy 534 AATATGAAATCTTACCTCTCAGAGGTTTCTTCTTACTGAGTTTGAATCTCTCTAGAC 593
Db 354 TTTCTGGAATTCAGAAATTCAGAGATTTTCAACCTTGGATGCTTGCAGCTCTAG--- 410
Oy 594 AACGTTTACTACAGTCAAGGACCTTCTCAGACAGAGGCGTAATTTAAGAAATAT 653
Db 411 -----GACCAAGGTAAGGAGGACCATTAATTAACGATGCTGTGTTGAGAGTAT 464
Oy 654 TCGTAACCTTGTAGTGTGGGAATTTTCTACTGCAAGATGGTGAGCTATCAAGAGAC 713
Db 465 AGGGAATTTTGACCAAAATGAAATGCTCTAGTGAATGAGGAGGACCATCAATAGAA 524
Oy 714 GTCTTCTTTTAACTGCACTTGTGAGATGCTCTTTTAACTTAACTTCAATCAAC 773
Db 525 GACTTGTCTTGAAGTGGAGTACGCGGTTTGTAGCGTTTGGCAATATGCTGTCGA 584
Oy 774 AAAGGAGAGCAATTTGCTACTACAGAGGCGTCGATAGCAATTAACAGAGTTATGT 833
Db 585 ACAGGGGAGAGATCTATGCTTGTGAGTCTGTGATTTCTGGAATGCGAGATCTT 644
```

```
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB6 protein"
/protein_id="CA37081.1"
/db_xref="GI:445897"
/db_xref="SPTREMBL:Q92382"
/translation="MTAAPALHVSPTPEETKEFVLPKDSKRALGILLVGLV
CAIVLSGIVSGIALIVGIGISTSLGVLPVLGILLIRRELTLREKQIAE
TPADELKEIMYIOSTERSLEKIEGSRYSDDGLNPNATOKIIDLESLSITSEFDL
KOLFDEKTELISGERLLEFLIANLFRKGRDYLNLGNLADIRAYGMNNYKAMVIE
KAKAVHEIFIVITMARLEFF"
13014..14585
/gene="hb7"
/size="14585"
/contig="14585"
/cds_start=1
/codon_start=1
/transl_table=11
/product="HB7 protein"
/protein_id="CA37082.1"
/db_xref="GI:445898"
/db_xref="SPTREMBL:Q956P3"
/translation="MYQENIRLERLLNSVOKSYADRLFSYEKTKVHDPLPWE
DEKCAEAKAPLEOKILIDYKSIFFMLNENDEIMLPMSGLTVPTRKFOED
DESRMHKVLIOKLEDNYEKLLEESKESEENANKILSDVLDLDAKAPFLKOE
VERVYKDLARAGYGVDPKODPEAKKKVLEASLEFLDISSELYOCLEDODIYKE
QVYKDLARTQOELEBDIEAKREBAADLSLNERLKSRTMDRAKHNENADSTW
WTSQLEMKDKAKALKLLEDITSVPEIDIEFCSLELPLTLTRELKSLYKRI
CSETLKMYSVENNNITYVOEYVONLONLQKLOGISQREKQKODDPAENLQVLAOK
RLRELTONFEIOGFNFKEDFKAAADLYIRSTAEQKMFVDVCMELFRYHEVNR
LLEMYNCADSYRDAKKLCSLRDEKELLQRIKKEEYOKKQGRHADRSHRTYQK
LRIAELELEIKKI"
14954..17479
/gene="omp7"
/size="14954"
/contig="14954"
/cds_start=1
/codon_start=1
/transl_table=11
/product="omp7"
/protein_id="14954"
/db_xref="omp7"
/translation="omp7"
/size="omp7"
/contig="omp7"
/codon_start=1
/codon_start=1

Query Match      15.6%; Score 498.8; DB 2: Length 26920;
Best Local Similarity 51.6%; Pred. No. 9.3e-95;
Matches 1395; Conservative 0; Mismatches 1237; Indels 74; Gaps 8;
```

```
Db 21310 GACTTGTCTTTCGACGTGGAGTACGGCGTTGTAGCTCTCTTGGCAATAGCTGTGCA 21369
QY 774 AAGGAGGAGCAATTCTCTACAGCAGCGCTCGATAGCAATTAACAGATTATGT 833
Db 21370 ACAAGGGGGGCGATCTATCTCTGTGTGACTGTGTATTCGAGAAATGCAATCTT 21429
QY 834 TAGATTCCTATCAATACATACGCTTACGTAGAGAGCGGTATGATGATGAAGGACGTC 893
Db 21430 GAGCTTGGGAACAACAGTGCACACATCAGAGAGCGCATCTCTCTGAAGGAACCT 21489
QY 894 GATATTCGACAAACAATTTCTATATTTGAAGGAATGACCGAAACTACTGCGG 953
Db 21490 TGTGATCTCCAAATACCAAAATATCTTTTGTGATGCTGCAAGACACTCAAAATGCGG 21549
QY 954 TGGCAT-----CTGCAACACCAAGGCGATGCTCTCTGAACTGTATTCCTTAACA 1007
Db 21550 AGCTATTCGATTAACCAACAGAGGCGAAACCGACACTATCTTGTGACTTTCAGGAA 21609
QY 1008 TAAAGACTGTATCTTCTTCTTCAAACTGACGAAACAAAGCGTGGCGCATTCATGCTAA 1067
Db 21610 TGAAGCGCTGATTTTCTGATTAACACAGAGAAATAGTAGGTGCGATTTATACCA 21669
QY 1068 AAAAGCTAGCCCTTCTCTGAGGCTTTACAGATTCTACGAAATATGTCATC-- 1124
Db 21670 AAAATGCTGTATTCCTCAGAGAGAGAGAGTGTATTTTCAACAAACAGCTGCGAA 21729
QY 1125 AGCACTCTCTAAGGGGGTCTATCAGATCGATGCTCAGAGAGCTCATGCTTTGTC 1184
Db 21730 TGTATCTCCCAAGAGAGGCGCAATTCGATTCAGATTCGAGAGATGATTCCTG 21789
QY 1185 AGAGACAGGAACATTCCTTGTGAAGAATACCTTACAAACCGAGTACCGATAC 1244
Db 21790 AGATCTCGCAATATCTATTTTCGAGGCAATACAGAGACCTACAGAGAGTCTGAG 21849
QY 1245 TCTTAAACGTATTCGATCAATACATAGAGAAATACGGAATTCAGGAAATTCAGGCTGC 1304
Db 21850 TGTGACCAAGAAATGCTATGATCTTCGATGCAATGCAAAATTTTAAATCTCGAGGAC 21909
QY 1305 TAAATATCATCAATTTCTCTATATGATCCCATCAGCTTCAAGAAACCTCATCAGCT 1364
Db 21910 TCGGGGAATTAAGTATTTCTATGATCTTATCAGAGCTCAGG--GCTACTGATTA 21966
QY 1365 ATGAAGATTAATACGCGCTCTCGGAGCTCTCAATCCATCAACAGAGATTCAT 1424
Db 21967 GCTCTCTTGAATTAAGCTACGAGAGATTCGAAATACCTATGAAGCTACATCGTTT 22026
QY 1425 TCTTGAAGAAACCTTAACAGAGATGAATTAAGTTGCTGACAAATTAATCTTCATT 1484
Db 22027 CTCTGAGAGAAACCTCAAGAGAGAACTTAAGAAACCTGACAAATCTGAACTCATATT 22086
QY 1485 CAGGACCGAGTCCCTTATCCGAGGAAAGTTATTCCTCAAAAGAGTCACTTAA 1544
Db 22087 TACACAGGCTGTAGAGTGTCTGAGGTGCTTATGATTAAGATGAGTGTGAT 22146
QY 1545 GAGCAGAGGCTCTCTCAAGAGCGGCTCTCTCTGCGATGATTCAGAGACGAT 1604
Db 22147 TGAATATACATTAAGAGAGTTCAGAGATGGAAGTGTATGATGAGAGGACTCTT 22206
QY 1605 ATCAATACAGCTGGAGATTAATCAATCAGAACTTAGAATCAATGTTGATCTTAG 1664
Db 22207 TGAAGCAAGGCTGAGGGGCTCACTCAATGCTTACGATTAATATATATATCTTAGA 22266
QY 1665 TCTTAAGCAGCCCTGACGCTTAACAGCAAGAAAGTGTCTCAATTAAGTGTCTATCTG 1724
Db 22267 TGGGACAAATTAAGCTATCTTAAGGAGGAGGAGCAAGTAAGATGATTCCTTATAG 22326
QY 1725 GAAGCTCAACCTGATGATTTGAAGGAACTTTTGAAGGATCATATGTTCAAGCATGA 1784
Db 22327 GCTTATCATGCTGTGATGCTGAGGAGACTTTATGACATCATATCTCAGTACACA 22386
QY 1785 CCAAGCTCTCTCTTATTAATAATCAAGGTTGATGCTGATGTTGATACATGATGAT 1844
```

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 26920)  
AUTHORS Boesen, T.  
TITLE Direct Submission  
JOURNAL Submitted (21-JAN-1999) Boesen T., Department of Medical  
Microbiology and Immunology, University of Aarhus, The Bartholin  
Building, DK-8000 Aarhus, DENMARK  
Location/Qualifiers

FEATURES  
source 1. 26920  
/organism="Chlamydomophila pneumoniae"  
/strain="VR1310"  
/db\_xref="taxon:83558"  
498. .3266  
/gene="omp6"  
498. .575  
/gene="omp6"  
498. .3266  
/gene="omp6"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein 6"  
/protein\_id="CAB37075.1"  
/db\_xref="GI:4453891"  
/db\_xref="SPTREMBL:Q92449"  
/translation="MRFSLCGPPLVFSLTLLSVFDTLSATTISLTPEDSFHGDSQNA  
EESYVQAGDVYSLTGVDYISNVDNSALNKACFNTSGSVTPAGNHGGLFENNISGT  
TREGAVICCCDPOATARFSGPSTLSFIOGDIKROGCLYSKNALMLNLYNVPREON  
QSKTRGCAISGANTVIVGNIDSVSFYQNAATFGAIIHSGPLQIIVNQAETPFAONTA  
KNSGGALYSDDIDIDQNAVYLFRENEALTTAIGKGAVCCLPTSGSSTPVIPTFS  
DNKQLFERNHSMIGGAIYARKLSISSGGPLFINNISYANSONAGAIADTGEI  
LSAEGKGTTEGQNTSLPLNGIHLQNAKFLQARNGCSIEFYDEITSDAGSTQ  
LINRDPKNKEEYGTILFSGEKSLANDPDEKSTIPONVNLGAGLYIKEGAEVYVK  
FQSGPSHLVIDLGTKLJASKEDIAITGIAIDISLSSSTAAVITKANTANKQISLT  
STELISPGNAEYEDLRMNSOTFPLSLPRGAGSVYTAGDPLVSPHIGQGWKL  
AMTGINKVGEFTFDKINIKPRPEKGNLVPIYLNNEITPKHYTSMASQLEFSKQDY  
RGLWIDGIGNFHVASSEDNIRYRHNSGCVYLSVNNETPKHYTSMASQLEFSKQDY  
AVSNNEYRMVLYGSLYQYTTSLNGINFRYASRNPNVNGILSRFLQNPMLTFHLCAY  
GHATNDKMTDYNFPMVKNNSMNCMAIECGGMPFLVFNENGLFOGAIPEMKQLQVY  
AYOGDFETTAGDGRFNSGSLTISVPLGIREPKALISODVLYDPSFSYIPDPEKDP  
SCSEALVTSQDSWLVPAHVSNAFVSGCTGRHYHNDYTELLCRSIECRPHARYNI  
NCGSKFR"

gene 3976. .6870  
/gene="hb1"  
3976. .6870  
/gene="hb1"  
/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB1 protein"  
/protein\_id="CAB37076.1"  
/db\_xref="GI:445892"  
/db\_xref="SPTREMBL:Q923C8"  
/translation="MRKRLIATVILAIISIIAGCVLLVYAIPLGSSVISPAGMG  
ACALCGVMAIDIVLKRREPIVLASVTTPTGGSPRSGISISGASAKMLIHGMPF  
LDEGHPOSMRKRLIATVILVFSIIILASGVLLVYAIPLGSSVISPAGMGACALG  
VMAIAGIDVLLKREVPVILASVTTPTGGSPRSGISISGADSTIRSLPTPIEDGH  
OSMRKRLIATVILVFSIIILASGVLLVYAIPLGSSVISPAGMGACALGCVMAIG  
IDVILKREVPVILASVTTPTGGSPRSGISISGADSTIRSLPTPIEDGH  
VESLNNKSLPYDGHGLEBKTKHOIRYRSSIKAMPFELDIRIFEEBEFFTSAR  
KRLDLATTLVERKLTLEQLERNNLKAFSVLYQDSIFPKIINDFEKIAMFMILSKS  
ICRFTIIPENHEHGAVALHNAVLLEKVIYRSLOKSYRDIKMSAKMLIHGMPF  
SLEDKKTIMKEHGAVALHNAVLLEKVIYRSLOKSYRDIKMSAKMLIHGMPF  
ISRDQOMKRAHLKHOESLIYQARDRLTQDSKNOKELEKAEQYIISWBEVKKEFI  
ERVORIRAIQIKYINILEREETGOETVPTVGTQASSDIIILIRIYSSSEDN  
ONQESCVYLVASHEVEMSWYKQETGPKKKEFQDQMGSLERFTTHIELEVLQDYS  
KHLSTFKVNNKKEVQAKFRKLVESDLEGLAQTESAESLTQEELPTILATRCLE  
KAVFGSLCCALASAKAPYFEDPFRQSDTQLRALTLRLQAKASLEELIKRFSNLE  
NDYIEERLLEKQTFERAGIYREIAVESTYLRSLTJTWESTPSESEKYVYSMTL  
NYVNEERAKRKLIVEMTQRYRDFKMALEAQNFEALQELSLIQAPSE"

gene 7150. .8058  
/gene="hb2"  
7150. .8058  
/gene="hb2"

CDS

/gene="hb2"  
/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB2 protein"  
/protein\_id="CAB37077.1"  
/db\_xref="GI:445893"  
/db\_xref="SPTREMBL:Q925Q3"  
/translation="MQVFLSPOLPPPPRSIGVSGSKRLSISTLLVGLVLLTGLIP  
GLTAGISGAGCGFSALCGVLYISGLLFLVRRPVPYRSEIPGSGVSGSEPALE  
KAQKEPRTKLTIDRLPKELDDLDITYIQEVFKLEMLKDPKEDRELLEAKRKLVFD  
VVEKDMSEFLDIQVRLNEEAIVYEHCDPLEINAYELFSSQELNDYVCAQCGPLPS  
GDARADRLKRSYKEVMDRFRMYTMSWEASVLDHSYGVARLELFAVGVLEESYKI  
LFKSYRADYFEECEKAIQDQGFKWL"  
8185. .9492  
/gene="hb3"  
8185. .9492  
/gene="hb3"  
/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB3 protein"  
/protein\_id="CAB37078.1"  
/db\_xref="GI:445894"  
/db\_xref="SPTREMBL:Q923C0"  
/translation="MMLDRYADKFLIRKEEKEMERHELHATMVRKASGHAYAKAA  
FEKERSNORKVKQDEKMKLSGLAEFRNOSRRARERLRELOTLYPEVSEYELER  
ORTKVLNLENLYADIEKKYHCVRQDEHYMKVEVKEAEYRENGKVSAAEVSQIC  
RLDQCLVNLSEKTLTAESVFEBAKDATEKLGKLNKYLSPYTNRELICLDAEEMTRIE  
EIEMLTLNVELPLLEKNTPEKASIQVNSCKEMLAKVPQCKESFTYSSQERELRN  
QDLQVAVNCCORLEQFSDLESKYTCRDLREOKAHEVQGLNFINEELVWGLFL  
TOARDIATVAYMEFYQVYHNIRKEKVSQOMMAKTERYREIROAFQGVKREDLIAED  
TIKEDYDWLRLDMLLDERKNRQRLICNKIAAQQRVKGF"  
9805. .10698  
/gene="hb4"  
9805. .10698  
/gene="hb4"  
/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB4 protein"  
/protein\_id="CAB37079.1"  
/db\_xref="GI:445895"  
/db\_xref="SPTREMBL:Q923B1"  
/translation="MLLISGALFYLIGIPGLSAAISFGLIGISALGVLMSGLIC  
LLVKEIPIVAREEIPBEGVSLAPSEEPALQAAQKTLAQPKELDQLDNDIOVFACIR  
KLKDSKYSRSTLNDARKKELVDFDVVEDTISEIFELAQIYVAQBECDMLNPLNGRSL  
MMTASESLDLFHVSKRLGYLPSGDVREGLKSAKETVARLMSLHCHLHVAAVAFDR  
NSYAAAEKAFKALGALDESYSRLTQSYRDKFLSEERAKIIPWNGHITWLDDAKSGC  
AEKRLGMPRNVRNLRNGKQSGF"  
10776. .12119  
/gene="hb5"  
10776. .12119  
/gene="hb5"  
/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB5 protein"  
/protein\_id="CAB37080.1"  
/db\_xref="GI:445896"  
/db\_xref="SPTREMBL:Q929G3"  
/translation="MDEITFHELYEKTFFLRLHKKALAKTTEKKRSKRNLOAVEE  
ANARLKVARMYDOEFQKAGRELEKTLALYPEVYSIRENKIOETRSNLEKAYAE  
ENYRCVREOQDQWKEEERKREAREPERENKTLISPEKSSLEOFDGHGKNSRSEAL  
EGHILKLOKEMTAVYENKTLIDABSRLEIYVEDYEMPCRIIELEKTLRMALSTPLPT  
KKAERKACSQYNSCAEMLEKYPYCKESLAVYTSKERLVSLDEDLRRATYRCQKRFQ  
DSGLESEVRACREDUERRIOEFETQGLDVERKELLCVSSRLNTECDVSGVKKEAAP  
GKKFYAQYDEIYRVRQSMWTMSERLREGVQCNKMLKAGLSEEDKVLKEEYWL  
REERNKKEKRLVGTIVATQRIQIEFQSDIYESNKEVSLMDKARFLFNEDHS"  
12262. .12987  
/gene="hb6"  
12262. .12987  
/gene="hb6"

CDS

QY 1125 AGCACTCTTAAGGGGGTGGCTATCAGATCGATCCCTCAGAGAGGCTCAGTCTTTCGC 1184  
 Db 4140 TGGTACCTCCTTAAGAGGGGCAATGGGATTCCTATGATTCGAGAGATAGATTTCTGC 4199  
 QY 1185 AGAGACAGGAAGAACTTACCTTTGTAAGAAATACCTTTACAAACCGAGTACCGATAC 1244  
 Db 4200 AGATCTGGGCAATATCTTTTTCGAGGCAATACAGAGCTACAGAGAAAGTCTGCGAG 4259  
 QY 1245 TCCTAAGCTAATGCGATCAATAGAGAGTACGGAATTCAGGAATTAAGGGCTGC 1304  
 Db 4260 TGTGACCAAGAAATCTATAGATCTTGATCGATGCAAAATTTTAAATCTCGAGGAC 4319  
 QY 1305 TAAATAATCATCAATTTTCTATGATCCCATCATCTTACAGAAACCTCATGAGAGT 1364  
 Db 4320 TCGGGGAATTAATGATTTTCTATGATCCGATCAGAGCTCAGAG --- GGTACTGATTA 4376  
 QY 1365 ATTGAATTAATTAAGGCTCTCGGAGAGCTCTCATTCATATCAGAGAACGATTTGAT 1424  
 Db 4377 GCTCTCTTGAATTAAGCTGACGAGAGATCTGAAATACCATGAAAGCTCATCGTTT 4436  
 QY 1425 TTTGGAGAAACCTTAACAGAGATGAACTTAAGTTGCTACAAATTTAAATCTCAT 1484  
 Db 4437 CTCTGGAGAGAACTCTCAGAGAGAACTTAAGAACTCAGAACTGAAATCTCATATT 1496  
 QY 1485 CACGACACCAAGTCTCCCTATCCGAGAGAAAGTTATGCTACAAAAGGAGTCACTTGA 1544  
 Db 4497 TACACAGAGCTGTAGAGTGTGTCGAGGTGCTTGTATGTAATGAAATATGAGAGATG 4556  
 QY 1545 GAGCAGACGCTTCTCTAAGAGGCGGCTCTCTCTGCGCATGATTCAGAGACGATTT 1604  
 Db 4557 TGCATAATCAATTAAGAGAGTGCAGAGATGAAAGTGTATGATGAGAGGATCTATT 4616  
 QY 1605 ATCAATACAGCTGGGAGTATTAACATCAGAACCTTACGATCATGATGCTCTAGG 1664  
 Db 4617 TGAAGCAGGCGTGAAGGGGCTCACTCTCAATGCGTACGCTTATATAGTCTTTGA 4676  
 QY 1665 TCTTAACAGCCCTCAGCTCAGCTAAGCAAAAGTCTTCAATTAAGTATGATCTGATGG 1724  
 Db 4677 TGGGACAAATTAAGCTATCATTAAGGCGAGCGAGCAAGTAAGATGTGCTTATCAGG 4736  
 QY 1725 GAGCTCAACCTGATTAATTAAGGAACTTAATGAAGATCATATGTTGACGCCATGA 1784  
 Db 4737 GCTATATATCTGTGATGCTGACGAGAACTTAATGAGATCATATCTCACTCAACA 4796  
 QY 1785 CCAAGCTCTCTCTATTAATAAATCAAGGTTGATGCTGATGATTAATCACTCACTGAT 1844  
 Db 4797 GCAAGTCTCTCTTAATAGAGCTTCTGCAAGAAAGATGATCACTCACTGATATCCC 4856  
 QY 1845 CAGCAGCTTATCCCTGCTGCTGAGATCTTAATTCAGAAATACGATTCAGAGCA 1904  
 Db 4857 CGATACCCCAATCTAATTAATCTAGATCACTAT-----GGGTATCAAGGAGAC 4905  
 QY 1905 ATGGAATGTTAATTTGAGCTAGCTATCAGCTACAAATACAAAGAGCCAGGCACTTG 1964  
 Db 4906 TGGAAATTAATTTGGGTGACAGATGCACTGCAAAAACAAAATATCTACCTTAACCTG 4965  
 QY 1965 GACCAAAAGAGATTTGCTCCAGCCCGGAAAGAAATGCGCTTATATGACATACCT 2024  
 Db 4966 GACTTAACAGAGATACAGCGAATCCAGAGAGCTTGTGCTTCTTAATACCT 5025  
 QY 2025 ATGGAGAGCTTATCTAGATTCGCTCTGCAACACCTTGTAGAGATCGGCGCACTGG 2084  
 Db 5026 GTGGGGTCTTTTGTGATCTCGGCTCATATGAGGCTCATGAGCGGAGCAACATTTTC 5085  
 QY 2085 TATGGAACACAAACAGGTTTCTGGGTTTCTCCATGACGAACTTCTGCTAATAGACTGG 2144  
 Db 5086 GTTATCTCTCTCAACAATTTGTGGGTATCAGAGATCGCGAGCTTTTGCATGAAATCA 5145  
 QY 2145 AGATGAATAATCGAAGGCTTCGCTATCTCTGAGAGCTACTGCTATCGGTGGAATGCG 2204  
 Db 5146 GAAAGGAACCAACGATGATCTCATCTTACGCGGGTTATGCAATTAAGAGAGATTT 5205  
 QY 2205 TCACACTCTAAGAGACGATTAATTAATCTTGGGTTCTGCCATCTCTTCTAGAGACAA 2264

Db 5206 CTTCAGGAGCTCTGAAATATTTCTTAATTTTGTGCTTTTGTGACGCTTTTGGCTACGACAA 5265  
 QY 2265 AGATTTGTTTATGCTCTACAAACACTTAGAACCTACGGTGGAACTTATTTCTCAAGCA 2324  
 Db 5266 GGACCATCTTGTGCTTAAGAACCATTAACCATGTTATGACGGGGCAATGAGTTACGACA 5325  
 QY 2325 CTCTCATACCTTACCAACCCCAAACTATTGAGATTAGAAAGACAAAGTTTCTGATC 2384  
 Db 5326 C-----CTGGAGAGCTTAAGACCTCGCTAA 5382  
 QY 2385 AGCTATGAAAAATTTCCCTAGGAAATTTCCCTAGATGATGATGATGATGATGATGATGAT 2444  
 Db 5383 GATTTTGTGAGAAATTTCTGACCTTACCTTGTGCTTCAATCTCGGTTTGTGCTATG 5412  
 QY 2445 CCATTCAGACACCGTATGGAAGACGACTATACCTCATTTGCGCAATCCAGAGTTCTTG 2504  
 Db 5413 CCATACCGCAATTAACATGACCAAAAGTACACTGAGCTATCTCTCTGTTAAGGAGCTG 5472  
 QY 2505 GAGCAACGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2564  
 Db 5473 GGGAAATGATGCTTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5532  
 QY 2565 ACATCTCTTTTCAAGACCTTATTCACAGATGAAAGTCAAGTATGATGATGATGATGATGATGAT 2624  
 Db 5533 TCGGCTTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5592  
 QY 2625 AATATGCTTCTTCAAGACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2684  
 Db 5593 GAATGATTTAAGGAAAGGACAGACAGAGGCGCTTCTTCCAAAGTGAAGACCTTTCA 5652  
 QY 2685 CCTCTGATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2744  
 Db 5653 TCTAGCGGTTCTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5706  
 QY 2745 TGATCTCTCAAGATTTCTTGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2804  
 Db 5707 TGATCTCTCAATGATTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5766  
 QY 2805 TCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2864  
 Db 5767 TCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5826  
 QY 2865 TTTATGAGGAGTGAACAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2924  
 Db 5827 TCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5886  
 QY 2925 CGTATGGAACCTCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2984  
 Db 5887 TGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5946  
 QY 2985 ATTCTA 2990  
 Db 5947 ATTCTA 5952

RESULT 11  
 CPN133035  
 LOCUS CPN133035 26920 bp DNA BCT 27-FEB-2001  
 DEFINITION Chlamydia pneumoniae GAI-protein family cluster B, genes omp6,  
 hbl1, hb2, hb3, hb4, hb5, hb6, hb7, omp7, pseudogenes omp8, omp9,  
 omp15, strain VR1310.  
 ACCESSION AJ133035  
 VERSION AJ133035.1 GI:4455890  
 KEYWORDS HBI gene; HB2 gene; hb3 gene; hb4 gene; hb5 gene; hb6 gene; hb7  
 gene; omp15 gene; omp6 gene; omp7 gene; omp8 gene; omp9 gene; outer  
 membrane protein 6; outer membrane protein 7; pseudogene.  
 SOURCE Chlamydia pneumoniae.  
 ORGANISM Chlamydia pneumoniae.  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 REFERENCE 1 (bases 1 to 26920)  
 Daugard, L., Hjerno, K., Knudsen, K., Madsen, A.S., Christiansen, G.  
 and Birkelund, S.





QY	1806	AATACAGTTTCATCTGATGTGTGATACTACATCGTTCACATACAGACCTTATCCCTGTCC	1806
Db	1638	ATTCC - - TCTCTTACTCTCTGTCTACCGGATTTTATATGATGCGCTTCTACTCTTCC	1694
QY	1866	TGCTGAGGATCTCTTAATTCAGATAATACGATTTCCAGAGACATAGGAATGTAAATTGGACTAC	1925
Db	1695	AGTACAAACTCCACAACTCTATTACGGGTATCCAGGACATTGGGAAGCCACTTGGGCGA	1754
QY	1926	GGATTCAGCTTACATACAAAGAGCCACGCGCACTGCGACCAAAACAGATTTGTTC	1985
Db	1755	CACATCAACTCCAAAATCA - - - - - GGAACATTATGATTTGGTAACTACGCGCTACACACC	1808
QY	1986	CAGCCCGGAAGAAATCTCGCGTTAGATATGCAATACCCATATGGGAGTCTTACTGACAT	2045
Db	1809	TAAATCTGAGGCTAGACCTTCCGTAGTCCCGATTCTATTATGGGATCCCTTATCGACAT	1868
QY	2046	TCCGCTCTGCAACAGGTTGTAGAGATCGGCGCACTGGATAGGAACAACAAGGTTT	2105
Db	1869	TCCGACTCTTCAGACAGATCATGACATCTCAAGGAAATATATCTATACACAGAGGACT	1928
QY	2106	CTGGGTTTCCCTCATGACGAATCTCTGCATAGACTGAGATGAAATCGCAAAAGGCTT	2155
Db	1929	CTGGGCACTACAGGAACCTCGAATTTCTTCCATATAGATTAATCAGGAACGTACCAAGACT	1988
QY	2166	CCGTATACCTCTGGAGGCTACGTCAATCGGTGGAGTGTCTCACTCTTAAAGACACT	2235
Db	1989	CCGACATATAAGCTACGCGCTATATGTTGGAGGAACTGTGAAGATTTTCTGAAATAT	2048
QY	2226	ATTATACCTTCCGCTTCGCACTCTCTTGGTAGAGCAAAAGATGTTTATTCGCTCAAA	2285
Db	2049	CTTCACTGCTACTTCTTCCACAGCTCTTCCGTAAGATTAAGACCTGTTTATATGTTGAAA	2108
QY	2286	CAACTCTAGAACCTACCGCTGGAACCTTATCTTCAAGCACTCTCACTACCTCAACCCCA	2345
Db	2109	TACCTCTCATATCAATTTAGCGTGGCATATCTCGCAACATCGAGACTTCTTAGAGACT	2188
QY	2346	AAACTATTTGAGATTAGGAAGACAAAGTTTCTGATATGACTATAGAAAATTCCTAG	2405
Db	2169	TCCCATGGCCCTATTTTGAAG - - - - - TATCACTCGACATGCTCAA	2207
QY	2406	GGAATTTCCCTTACCCTTGGATGTGCCAAGTTTCAGCCATTCAGACCAACCCGATGA	2455
Db	2208	AGATATTCCTCTCAATTTTGAATGCCCCAGCTAAGCTACAGCTACTAAAAATGATATGA	2267
QY	2466	AAGCATTATACCTCATTTGCCAGAAATCGAAGTTCTTGGACCAACGAGTATATACGG	2525
Db	2268	TACTCGTATACCTCTCATCTGAGCTCAAGTCAAGTTCTTGAGCATAATATTCGGGGCTCT	2327
QY	2526	TGCTATGGGCTTAGACCTTCTCTTTGTTCTTTCCAACCACTCTCTTTCAGACCTT	2585
Db	2328	AGACCTCGGAGAGATCTGTGCTCTATATCTCCCTTAAAGACCAACCGTTCTTCAGGGATA	2387
QY	2586	CATTCCACAGATGGAATGCAAAATGTTTTATGATACCAAAATFACCTTCTCGAAGCTC	2645
Db	2388	TTTTCCCTCTTAAAGTTCCAGGCAAGCTTACACCGCCCAACAAACTTTAAAGAAAGTGG	2447
QY	2646	TAGTATGAGGCCGTGGTTTGTATGATTTGGAAGGCTGTCTTCACTCTCGATTTCCGTGGGTGC	2705
Db	2448	CGCTGAAGCCCGTGTCTTATATATGAGACCTCAGTGAAGCTCTCTATCCCTCGTGGCAT	2507
QY	2706	GAAATTCGTGACGGGGGATATCGGAGATTCCTAACACTATGATCTCTCAGGATTCCTTGT	2765
Db	2508	TCCGTTATGAAAAAATCTCCCAAGATGAAAAAATAATTCGAGATTTCTTACAGCAACAT	2567
QY	2766	TTTCCGATGTCTATACGTACAAATCCCAATCTACAGGACTCTTGTGATAGAGCCACGATC	2825
Db	2568	TGTGTATGTATATCGTAAAAATCCCGTTCTGCGTACTTCTTAAATGGTCACTGGAGCTC	2627
QY	2826	TTTGAAATTTCCGGGTGGCAATCTTTCACAGACAGGACTTTTACTGAGGGGTAGCAACA	2885
Db	2628	TTTGACTCTGCTATGTAAAAAAGCTCGCAGACAGACAGCTTCTTAGCAAGTGTGGAAGCA	2687
QY	2886	CTACGCTTACACTCCAAATTTGAGCTTTCGAGACTTACGCTATGAGACTCCGCGATC	2945

Db	2688	TCTAGCTCTCCCTCTACATGTAAGACTCTCTGCGAAGCTGCTTATGACCTTCGTGCTC	2747
Qy	2946	TTCAAGCACTACATGATAGATGTTGTAACCAAACTCCGATCTAG	2991
Db	2748	AGCACACATCTCATGTAGATGTGGGCTAAGATACATCATCTAG	2793
RESULT	10		
AE001587			
LOCUS	AE001587	16448 bp	DNA
DEFINITION	Chlamydia pneumoniae section 3 of 103 of the complete genome.		
ACCESSION	AE001587		
VERSION	AE001587.1		
KEYWORDS	GI:4376271		
SOURCE			
ORGANISM	Chlamydia pneumoniae CML029.		
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
AUTHORS	1 (bases 1 to 16448)		
TITLE	Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.		
JOURNAL	Comparative genomes of Chlamydia pneumoniae and C. trachomatis		
MEDLINE	Nat. Genet. 21 (4), 385-389 (1999)		
PUBMED	99206606		
REFERENCE	2 (bases 1 to 16448)		
AUTHORS	Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA		
FEATURES			
source	1..16448		
	/organism="Chlamydia pneumoniae CML029"		
	/strain="CML029"		
	/db_xref="taxon:115713"		
misc_feature	172..2957		
	/note="Polymorphic Outer Membrane Protein G Family: contains frame shift"		
gene	172..705		
	/gene="pmp_3.1"		
	/note="CPn0014"		
gene	615..2957		
	/gene="pmp_3.2"		
	/note="CPn0015"		
gene	3196..4971		
	/gene="pmp_4.1"		
	/note="CPn0016"		
misc_feature	3196..5953		
	/note="Polymorphic Outer Membrane Protein G Family: contains frame shift"		
gene	4874..5953		
	/gene="pmp_4.2"		
	/note="CPn0017"		
gene	6302..7786		
	/gene="pmp_5.1"		
	/note="CPn0018"		
misc_feature	6302..9139		
	/note="Polymorphic Outer Membrane Protein G Family: contains frame shift"		
gene	7787..9139		
	/gene="pmp_5.2"		
	/note="CPn0019"		
gene	complement(9380..11467)		
	/gene="CPn0020"		
CDS	complement(9380..11467)		
	/gene="CPn0020"		
	/codon_start=1		
	/transl_table=1		
	/product="Predicted OMP [leader peptide: outer membrane]"		
	/protein_id="AADI8173.1"		
	/db_xref="GI:4376272"		
	/translation="MKRCFLFLASFEVLGGSSADALTHQEAIVKKKNYSLSHFKSGVIV		



Db 14739 TGGTGAATGATGATGTAATAATCCCGTTGCGCTACTTCTCTAATGATGTCAGTGAGCCTTC 14798  
QY 2826 TTGGAAATATTCGGGTGGGCAATCTTTCAGACAGCGCATTTTCTAGGGGTAGCAACA 2885  
Db 14799 TTGGACTTTCGATGTAATAAACTCGACAGCAAGCCTTCTTAAGCAAGTCGGAAGCCA 14858  
QY 2886 CTACGCTCACTCACTCAATTTGTCAGCTCTTCGGACATTTACGATATGGAACCTCGTGATC 2945  
Db 14859 TCTGACTCTCTCCCTCCTCATGTAGAACTCTCTGGGAAAGCTGCTTAAGAGTGTGGCTTC 14918  
QY 2946 TTCAAGAACTACATATGATGTGTGTAACCAACTCCGATTCCTAGATTCCTAAACCTCC 3005  
Db 14919 AGCAGCACTCACTCAATGTATGTGGCTAAGATCTCATCTAGTCTCTACTTCTCTC 14978  
QY 3006 CTAGTCTCTCTAGGGAG 3022  
Db 14979 CCTAAACTTTAGGGAG 14995

RESULT 9  
LOCUS A81841 2793 bp DNA PAT 21-JAN-2000  
DEFINITION Sequence 15 from Patent WO9858953.  
ACCESSION A81841  
VERSION A81841.1 GI:6731871  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 2793)  
AUTHORS Madgen, A. and Birkelund, S.  
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE  
JOURNAL Patent: WO 9858953-A 15 30-DEC-1998;  
MADSEN ANNA SOPHIE (DK); BIRKELUND SVEND (DK)  
FEATURES  
source 1..2793  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 803 a 681 c 567 g 742 t  
ORIGIN

Query Match 17.2%; Score 551.2; DB 9; Length 2793;  
Best Local Similarity 52.7%; Pred. No. 8; 8e-106;  
Matches 1416; Conservative 0; Mismatches 1213; Indels 57; Gaps 8;

QY 330 AACGTTTACTCCAAAACCTTACGCCACACATATTTCTTAACAGAGAGTCTTCTTTTA 389  
Db 141 AATATCTACGCGATGCCAATGGAACGAACTATGTCCTATATCAGAAATGCTATATAA 200  
QY 390 CGAGCTGGAAGGACACCTCCCTTATCTGACATGTTGTTTAAACCAACCGACGAATCT 449  
Db 201 CGATGCTGGAAAGGACAGCATTTAACAGGCTCTGCTTTACGAAACTTACGGGATCT 260  
QY 450 TACCTTCTTGGGAGACGTCATAGCTTAACGTTTGGCTTATAGATCTGCACTCATGC 509  
Db 261 GACATTTACGGAAGGATACATCTTTCATTCACACAGGATGATGCGGGTTGCAATGC 320  
QY 510 AGGTGCTGCTGATTCACAAACGAATTAAGATCTTACTCTTCAGGGTTTCTTACT 569  
Db 321 AGAGGCTGGGAGCAAGCATGCTGATTAAGCCCTTAACATCAAGGATTTCTTAACCT 380  
QY 570 GAGTTTGAATCTCTCTACAGCACAAGGTTACTACAGGTCAAGGAAAGCTTCTCAGC 629  
Db 381 TTCTCTCATTTGAGCTCTCGGAACTACATGTTGTTACAGGAAAGATTTAGTTCTGC 440  
QY 630 AGAGGCTGAATTTAGAAATATTCGTAACCTTGTATGCTGGAAATTTTCTACTGC 689  
Db 441 AGAGGCTGAATTTAGAAATATTCGTAACCTTGTATGCTGGAAATTTTCTACTGC 500  
QY 690 AG-----ATGGTGGAGCTATCAAGAGAGCTTCTTCTTAACTGGCACTTCTGC 740  
Db 501 AGCTAATACATGCGGAGGATCAACCAAAAACCTCTTCTATTTCTGGGAATACCTC 560

QY 741 AGATGCTCTTTTATGTAACACTCTTCATCAACAAGGAGAGCAATGCTACTACAGC 800  
Db 561 TTCTATACCTTCTACTGATATATGCGCAAAAATTAAGTGGGCGCTATATAGCTCTGC 820  
QY 801 AGCGCTCGCATAGCAAAATTAACAGATTTATGTTAGATTCCTTCTTAACATATACCTCTAC 860  
Db 621 GCGTGAAGATTTTCAAGAAACCGCGCAGTTAGTCTTATGATTAATTAAGAGAAAC 880  
QY 861 GTACAGAGCGCTATGATGATGATGAAGCACGTGATATCTATCAACAACAATTTCTATA 920  
Db 681 TGGGGGCGGGGCTCTGGGCTTTGTAAGCCAGCTCTCGATTAACAAATGATCCCTTTT 740  
QY 921 TTTTGAAGGAATGACCGCAAACTACT-----GGCGGCGCATGCAACACAA 971  
Db 741 CTCTCTGGAACACTGTCACACATGCTGAGGCAAGGGGGGCCATTTATGTGAAA 800  
QY 972 GGGAGTGGATCTCTGAACGTAAATCTCTAACATTAAGACTGTGATCTTTCTTCAAA 1031  
Db 801 AACAGAGAGACTCTACTCTACTATCTCTGGAATTAAGTCTGACCTCGCCGAGAA 860  
QY 1032 CGTAGCAAGAACAGCGGTGGCCCATCTGCTAAAGCTAGCCCTTCTCTGGAG 1091  
Db 861 CTCTTAGTAACCTCAAGGGGAGCAATCTGTGCCATGTCTGATCTTCCCTGCTGG 920  
QY 1092 CTTTACAGAGTTCTTGAATAATGATCTCATGACAACTCT--AAGGGGGTGTAT 1148  
Db 921 CCTTACCCTATTTTCAATATATGATGCGGAAACACAGCTGCAAGGAGGGCGCTAT 980  
QY 1149 CAGCATGATGCTCTCAGAGAGCTCCTTCTGCAAGACAGAACTTACTTTGT 1208  
Db 981 TGAATTTGCGCATCTGAGATCTTAAATCTCTGCAAAATCAAGAGATCAAGTTCTCT 1040  
QY 1209 AAGAAATACCTTTAACAACACCGGAAGTACCGTACTCTTAACTGATGCTATCAAT 1268  
Db 1041 TGGCAACACTCTAACCTCAAC--TCCGGGCCAATCGACACAGGAATGCTATCACT 1097  
QY 1269 AGGAATACCGGAATTTACGGAATTTAGGGCTGCTAAATAATCATACATTTCTCTA 1328  
Db 1098 GGGATGCTCAGCAAAAATTAAGCACTTAAGGGCAGCCCAAGGCCATCTATATTCTA 1157  
QY 1329 TGATCCATCACTTTCAGAAAGAAC--CTCATCAGAGATTTAGATTAATTAAGGCTC 1385  
Db 1158 TGATCCGATTTGATTAACACACAGAGAGCTTGACAGTCTGACATCAACACCGGA 1217  
QY 1386 TCGGGAGCTCTAATTCATATCAAGGAAGATTTTCTTGGGAAACCTTAACAGC 1445  
Db 1218 TAGCAACTGCGCTTTAGATTTATTCAGGAAGATTTTCTGGGAAAGCTCTCTGC 1277  
QY 1446 AGATGAATTTAAAGTGTGACAAATTTAAATCTTATTCACGACGAGTCTCCCTATC 1505  
Db 1278 AGATGAAGGAAAGCTGTGATTAATCTTCACTATATTAACCAACCACTTGGCTCTAGC 1337  
QY 1506 CGAGGAAGATTTGCTACAAAAGGAGTCACTTAGAGAGCAGAGCTTCTCTCAAGA 1565  
Db 1338 CTCTGGAACCTTGCACTCAAAAGGAATGTCAGTTAGATGATCAAGTGTTCACACAGAC 1397  
QY 1566 GCGCGGTTCTCTCGGAGATGATTCAGGAGCAAGCATTAATCACTACAGCTGGAGTAT 1625  
Db 1398 TGAAGCTCTACCTCTCTACAGGAGCAAGCAAGCTCAAGCAAGATCTAAGCTAT 1457  
QY 1626 TACAAATCAGAACTTGAATCAATGTTGACTCTTAAAGTCTTAAAGCCGCTGACGCT 1685  
Db 1458 CAGTTTACCAAACTTGTGTGATCTTCTGCTTAAGAGGAATTAAGTGTGTCAT 1517  
QY 1686 AACAGCAAAAGTGTCTCAATTAAGTATGATCTGTGGAGAGCTCAACCTGATGATAT 1745  
Db 1518 TGAACAGCAGAGGCAACAAACTTAATCTTAACCTCTCTGTTTCCAAAGATAG 1577  
QY 1746 TGAAGGAACATTTATGAAGATCATATGTTTCAAGCAGACAGCTTCTCTATTAATA 1805  
Db 1578 TAGCGCAATTTTATGAAGGCAATACGATTAACCAAGCTTACAGCAGCTTGGTGGT 1637

QY	630	AGAGCGCTAAATTATGAAAATATTCGTAAACTGTGATGCTGGGAATTTTCTACTGC	689
D	12612	AGAGAGCTTAAATCTTACCGGAATAAGGAACGATTCCTTTACCAAAAGCTCCGATGA	12671
QY	690	AG-----ATGGGAGGCTATCAAAAGAGACGTCCTTTCCTTTAAGCGCACTCTGG	740
D	12672	ACCTATTAACAATATGGCGGAGCATTACCCAAAAACTCTTTATTTCTGGGAATACCTC	12733
QY	741	AGATGCTCTTTTATGTAACAACCTCTTCATCAACAAAGGAGAGCAATTCCTACTACAC	800
D	12732	TTCTATTAACCTTCACACTAGTAATAAGGCCAAAAAATAGTGGAGCGATCTATAGCTCGC	12791
QY	801	AGCGCTCGCATAGCAAAATAAACACAGGTTATGTAGATTCCTATCAACATAGCGCTAC	860
D	12792	GGCTCGAAATATTTTACGAAACACCGGCCAGTTAGCTCTTTAAGAAATAAAGAGAAAC	12851
QY	861	GTACGAGAGCGCTATCGATGATGAAGGACGTCGATACTATGAAACAACAAATTCCTATA	920
D	12852	TGGGGGTGGGGCTCTGGGCTTTGAAAGCCAGCTCTCGATTACTCAAAATAGCTCCCTTT	12911
QY	921	TTTTGAAGGAAATGCAGCGAAACTACT-----GGCGGTGGATCTGCACACCA	971
D	12912	CTTCTGTGAAACACTGCAACAGATGCTGCAGGACAAGGCGGGGCCATTTATGTGAAAA	12971
QY	972	GGCGAGTGGATCTCTGAACTGATATCTCTTAACATTAAGACTCTGATCTTTCCTCAA	1031
D	12972	AACAGAGAGAGACTCCCTACCTTACTATCTGTGAATAATAAATCTGACCTTGCCGAGAA	13031
QY	1032	CGTACAGAAACAAGGGGTGGGCCCATCCATCTATAAAAGCTAGGCGCTTCCCTGGAGG	1091
D	13032	CTCTTCAGTAATCTCAAGGCGGAGCAATCTGTGCCATGGTCTAGATCTTTCCGCTCTGG	13091
QY	1092	CTTTACAGAGTTTCTACGAATTAATGTCTCATCAGCAACTCCT---AAGGGGGTGGTAT	1148
D	13092	CCCTACCCATTTTCAAAATATAGATGGGGAACACAGCTGCAGGCAAGGGCGGCGTAT	13151
QY	1149	CAGCATCGATGGCTCAGGAGAGCTAGTCTTTCTGCAGAGACAGGAAACTTACTTGT	1208
D	13152	TGCAATTTGCCAGCTCTGGATCTTTAAGTCTCTGCAAAATCAAGAGGACATACGTTCTCT	13211
QY	1209	AAGAAATACCTTACAAACACGGAAGTACGCGATCTCTTAACGTAATGCGATCAACAT	1268
D	13212	TGGCAACACTTAACTCTCAAC---TCGCGCGCAACATGACACGGAATGCTATCTACT	13268
QY	1269	AGGAATACGGGAAATTCACGGAATTACGGGCTGCTAAAAATCATACAAATTTTCTCTA	1328
D	13269	GGGATCGTACGAAAAAATTAACAGCTTAAGGCGAGCCAAAGGCCAATCTATCTATTTCTA	13328
QY	1329	TGATCCCATCACTTCAGAGAGAAC--CTCATCAGACGTAATTGAAGATTAATACGGCTC	1385
D	13329	TGATCGATTTGATCTTAACACACAGAGACTTCAACGTTCTTGACCATTCACCAACCGGA	13388
QY	1386	TGCGGAGAGCTCATCACTCAATCAAGAACGATTTGATTTTGGAGAAACCTTAACAGC	1445
D	13389	TAGCAACTGCTTATGATTAATTCAGGAACGATTTGATTTTCTGGGAAAAGCTCTCTGC	13448
QY	1446	AGATGAACCTAAAGTTGCTGACAAATTTAAATCTTCATCTACGACGCAAGCTCTCCTATC	1505
D	13449	AGATGAAGGAAGAGCTGCGATTAATCTGACATCTATTAATTAAGCAACATTTGGCTTAGC	13508
QY	1506	CGGAGGAAGTTATTGCTACAAAGGAGGACTCTTTAGAGACGACGAGCTTCTCTCAGAA	1565
D	13509	CTCTCGAACCTTAGCACTCAAAAGGAATGTGAGTTAGATGTCAATGGTTTCCACACAGAC	13568
QY	1566	GGCGGGTCTCTCGCGCGATGAGATTTCAGGAACGACATTAATCACTACAGCTGGGAGTAT	1625
D	13569	TGAAGGCTCTACCTCTCATCAACACAGGAACAAAGCTCAAAAGACATCTGAAGCTAT	13628
QY	1626	TACATCAAGAACTTAGGAATCAATGTTGACTCTTAGGCTTTAAGCAGCCGCTCAGCT	1685
D	13629	CAGTCTTACCAACTTTCGTCGTTGATCTTTCTGCTTTAGAGGAAAAATGAAGTGTGTCCAT	13688
QY	1686	AACACAAAGGTGCTTCAAATTAAGTATGCTATCTGTGGAAGCTCAACCTATTTGATAT	1745

D	b	13689	TGAAAGAGAGGAGCCAAACAACTATACTCTAACCTCTCCTCTTGTCTTCCAAAGATAG	13748
O	y	1746	TGAAGGACATTTTATGAAGTCAATATGTTCACGCAATGACACGCTCTTCTCTATTA	1805
D	b	13749	TAGGGCAATTTTATGAAGCACTACGATTAACCAAGCCTTCACGACCTTGGTGT	13808
O	y	1806	AATCAGGTTGATGCTGANTGTGATACGTTGACATCAGACGCTTATCCCTGTCC	1865
D	b	13809	ATTCAAC---TGGTGGTACGCTGCTAGCCATATTTTATTCGATGCGCTTCTCACTTCTC	13865
O	y	1866	TGCTGGGATCTTAATTCGAATACGGAATTCGAAGCAATGGAATGTTAATGGACTAC	1925
D	b	13866	AGTACAACCTCCAGAACCTCATTTAGGGTATACAGGACATTGGGAAGCCACTTGGCCAG	13935
O	y	1926	GGATACAGCTACAAATACAAAAGAGCCACGGCAACTTGGACCAAAACGATTTTGTTC	1985
D	b	13926	CACATCACTGCAAATATCAGG------ACTATGACTTGGTAACTACGGGCTACAAACC	13979
O	y	1986	CAGCCCGGAAGAAATCTGCGTTAGTATGCAATACCTATGGGGAGTCTTACTGACAT	2045
D	b	13980	TATCTGTAGGCTAGAGCTTCCTCGTAGTTCCGATTCATATATATGGGATCCTTACTGACAT	14039
O	y	2046	TGCTCTCTGCAACGCTGTAGAGATCGGGCACTGGTATAGGAACCAACAGGTTT	2105
D	b	14040	TGCACTCTACAGCAAGATATGACATCTCAACGCAATATATCTATCACAACCAAGGACT	14099
O	y	2106	CTGGGTTTCTCCATACGCAACATCTCCGATTAAGACTGGAGATGAAATACGCAAGGCTT	2165
D	b	14100	CTGGGATACAGAACCTGCGAATTTCTTCATATAGGATTAATCAGAACATACCAAGCATT	14159
O	y	2166	CCGTCATACCTTGAGAGCTACGTCATCGGTGGAAGTGTCTCACTCCTTAAGACGACT	2225
D	b	14160	CCGACATTAACCACTACGCTATATTTGTTGAGGAAGTGTGTGAAGATTTTCTGAAATAT	14219
O	y	2226	ATTTACCTTGGGTTCTGGCAATCTGTTCCTATAGACAAGATTTGTTATCGCTACAA	2285
D	b	14220	CTTCACTGAGTCTTCTGCACTCTCTGGTAAAGTAAAGACTGTTATAGTTGAAA	14279
O	y	2286	CACCTCTAACAACCTACGGTGAAGATTTATCTTCAAGCACTCTCATACCTTCAACCCCA	2345
D	b	14280	TACTCTCTAATCTATTTAGCGTCCGCTATACCTGCACATCGAGATTCATGAGGACT	14339
O	y	2346	AAACTATTTGAGATTAGGAAGACAAAGTTTCTGAATCAGCTATAGAAAAATTCCTAG	2405
D	b	14340	TCCCATGCCCCATTTGGAAG-----TATCACCGAATCTGAA	14378
O	y	2406	GGAAATTOCCCTAGCCTTGGATGTCCAACTTCGTTCAAGCCATTCAGAACACCGATGGA	2465
D	b	14379	ACATATCTCTCATTTTGAAGCCACGATGAGCTACACATCACTTAATAAAGATATGGA	14438
O	y	2466	AACGCATTTACTCTCATTTGCCAGAAATCCAGAGTTCTTGGACACAAGATGATAGTG	2525
D	b	14439	TACTGCTATACTTCTATCTCGTAAGCTCAAGGCTTGGACCAATTAATCTGGGGCTCT	14498
O	y	2526	TGGTATCGGCTAGACCTCTCTTTGTTCTTCCAAACCCACATCCTCTTTCAGACCTT	2585
D	b	14499	AGAGCTCGGAGGATCTGTGGCTCTATATCTCCCTTAAGAAAGCACCGTTCTTCAAGGATA	14558
O	y	2586	CATTCCACGATGAAGTCGAATGTGTTATATATACAAATTAACCTTCTTGGAAGCTC	2645
D	b	14559	TTTCCCTCTTAAAGTTCCAGGACGTCACAGCCGCAACAAACTTAAAGAAAGTGG	14618
O	y	2646	TAGTATGGCCGTGGTTTATGATTGGAAGGCTGTGCTTAACCTCTCGATTCCTGGGTGC	2705
D	b	14619	CGCTAGAGCCCTGCTTTTATGATATGGAAGCTAGTGAATCTCTATTCCTCTGTGGCAT	14678
O	y	2706	GAAATTCGTGCAAGGGGATATCGGAATTCCTACACTATGATCTCTCAGATTCCTTGT	2765
D	b	14679	TGGGTAGAAAAAATCTCCGAGATGAAAAAATATTTTCGAGATTTCTCTAGCCTACAT	14738
O	y	2766	TTCCGATCTATGCTAACATATCCCAATCTACAGCGACTTGTGATGAGCCAGACTC	2825



Db	1143	CACATCAACTGCAGAAATCAGAGA-----ACTATGACTTTGGGTACTACGGGCTACAAACC	1090
Oy	1986	CAGCCCGGAGAAAGAAATCTGCGTTAGTAGCAATACCTTATGGGAGTCTTACTGACAT	2045
Db	1089	TAAATCTGAGGCTGAGACTTCCGAGTTCCCGATTCATTTATGGGCAATCCTTACTGACAT	1030
Oy	2046	TGCGTCTCTGCAACAGCTTGTAGAGATCGGCGCACTCGTATATGAAACACAAAGTTT	2105
Db	1029	TGCGAGCTCTACGAGAGATCATGACATCTCAAGCGAATATATCTATTCAGCAACGAGCACT	970
Oy	2106	CTGGGTTTCCCTCCATGACGAATCTCCGATTAAGATAGATGAGAAATCCGAAGGCTT	2165
Db	969	CTGGGCACTACGAGAACTCGAATTTCTTCCATTAAGATTAATAGGAACTAACCAAGCAAT	910
Oy	2166	CCGTCACTACCTTGGAGGCTACGCTCATTCGGTGGAACTGCTCACTCCTAAAGAGACT	2225
Db	909	CCGACATATAAAGCTAGGCGTATATTTGTGGAGGAAGTGTGAAGATTTTTCGAGAAATAT	850
Oy	2226	ATTTAACCTTTGCGTTCTGCGCATCTCTTTGCTAGAGACAAAGATTTGTTTATCGCTCAAA	2285
Db	849	CTTCAAGTGTAGCTTTCTGCGCAGCTCTCGGTAAAGATTAAGAACCTGTTTATGTTGAAAA	790
Oy	2286	CACACTAGAACCCACGCGTGGAACTTATTTCTCAAGCACTCTCATACCTTACAACCCCA	2345
Db	789	TACCTCTCATATCACTATTTAGCGTGCATATACCTGCAACATCGACATTCCTTAGGGAGCT	730
Oy	2346	AAACTATTTGAGATTAGGAAGACAAAGTTTCTGAAATGAGCTATAGAAAAATTCCTAG	2405
Db	729	TCCCATGCCCTCATTTGGAG-----TATCACGCAATCTGAA	691
Oy	2406	GGAATTTCCCTTACGCTTGGATGTCCAAAGTTTCTGACCACTTACAGAACCCGTATGGA	2465
Db	690	AGATATTTCTCTCATTTTGTGAATGCCACGCTAGACATCACTTACACTTAAATATGATATGGA	631
Oy	2466	AACGCACTATACCTATTTGGCAGAAATCCGAAGTTCTTGGACCAAGAGTGTATGCTGG	2525
Db	630	TACTCGCTATACCTTCATCTGTAAGCTCAAGGCTCTTGACCACTAATCTCGGGGCTCT	571
Oy	2526	TGATATCGGCGCTAGACCTTCTTTTGTCTTTCCAAACCCACATTCCTTTTCAAGACCTT	2585
Db	570	AGACCTCGAGAGATCTGTGCTCTATATCTCCCTTAAGAAGACCGTTCTTCCAGGGATA	511
Oy	2586	CATTCACACAGATGAAGTGCAAAATGGTTTATGTATTCACAAAATAGCTTCTTGGAAAGCTC	2645
Db	510	TTTCCCTTCTTAAAGTTTCCAGGCACTCTACAGCCGCCAAMAAACTTTAAAGAGAGTGG	451
Oy	2646	TAGTATGAGGCGCGGTTTATGTATTTGTAAGAGCGTATACCTCTCGATTTCCGTTGGGGTGC	2705
Db	450	CGCTGAAGCCCGTGTCTTTTGTATGTATGAAACCTAGTGAAGTGTCTTATCCCTGTGCGGAT	391
Oy	2706	GAAATTTGCTCAGGGGGATATTCGGAATTTCTACACATTTGATATCTCTAGATTTCTTGT	2765
Db	390	TGCGTTTGAAGAAATCTCCGAAATATGAAGAAATATATTTCCGAGATTTCTCTAGGCTATAT	331
Oy	2766	TTCCGATGTCTATGCTPAACATCCCAATCTACAGGAGCTTTGTGATGAGCCAGAGCTC	2825
Db	330	TGCTGATGTATGCTPAAGAAATCCCGCTTCCGCTACTTCTTAAGTATGCTAGTGAAGCTC	271
Oy	2826	TTGGAAATTTCCGGGTGGCAATCTTTCAGACAGGCAATTTTACTGAGGGGTAGCAACAA	2885
Db	270	TTGGACTTTCGCTATGTAAAAACCTCGCAGCAAGCAACCTTCTTAAAGATGTGTGAAGCA	211
Oy	2886	CTAGGTATCAACATTCGAATTTGAGCTCTTTCGGACATTTACGTTATGAGAAATCCGATGATC	2945
Db	210	TTGACATCTCTCCCTCTCATCTGTAGAACTCTCTGGGGAAGCTGTTTATGAGCTTCTGGGCTC	151
Oy	2946	TTCAAGGAACTACAATGTAGATGTTGTATACCAAACTCCGATTTCTAGATTTGCTTAAAACTCC	3005
Db	150	AGCACACATCTACAATGTAGATGTTGTGGGCTAGATATCATCTATGTTCTTACTTCTCTC	91
Oy	3006	CTAGTTCTTCTAGGAG 3022	

DB	90 CCTAAACTTTAGGAG 74
RESULT	8
AE001627	AE001627 15068 bp DNA BCT 01-DEC-2000
LOCUS	Chlamydia pneumoniae section 43 of 103 of the complete genome.
DEFINITION	AE001627 AE001363
ACCESSION	AE001627.1 GI:4376721
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	Chlamydia pneumoniae CWL029. Chlamydia pneumoniae CWL029 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE	1 (bases 1 to 15068) Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S. Comparative genomes of Chlamydia pneumoniae and C. trachomatis Nat. Genet. 21 (4), 385-389 (1999)
TITLE	99206606
JOURNAL	10192388
MEDLINE	2 (bases 1 to 15068) Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S. Direct submission Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES	Location/Qualifiers 1..15068 /organism="Chlamydia pneumoniae CWL029" /strain="CWL029" /db_xref="taxon:115713" 99..626 /gene="Cpn0439" 99..626 /gene="Cpn0439" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AADI8583.1" /db_xref="GI:4376722" /translation="MSSITGVNFPSSLPESADLFTTKKEIVALEGKNEVLTSHIP HIAITTIYVIALAGIATIGCYSSQSTILAVGVLTLLGLQALVGFKEFRQI PQGLHTYVIFREKIRPSSQLVTNMQRTQDTLKIYBELCDLSQKERNQSLTYK KRELSHNEKNTON" 651..1289 /gene="Cpn0440" 651..1289 /gene="Cpn0440" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AADI8584.1" /db_xref="GI:4376723" /translation="MATSVPSPYSPSSPSLHAEVNLTPAYITOPHPAPAEWET RSKSHRTLCFALTLLITDGCTISAGTACGTGMIIICIGLGIIVLLIALLATET LKNQGTGKRLIDEISODISSIGSFVRYGLMFTISVHLPELTNOQETRLNEE EAKESIONLEKLTTECONKLAQKPRKSSQKSPMSIKHLSKNPVILFDC" 1437..2387 /gene="Cpn0441" 1437..2387 /gene="Cpn0441" /codon_start=1 /transl_table=1 /product="CT007 hypothetical protein" /protein_id="AADI8585.1" /db_xref="GI:4376724" /translation="MFKLHITAPAGHVSTPIFYVODACIDBEACKNPPRPFS QVOVLNDAKFKRLPQITGROYDFTFLCTLPTTESGSLFTGYGADIQKSSSI PISTDDNGLMATFODTSEFNVLISGATLSLKNQWMSIILSGLDLPNIEKSSY LYQSLGCGVQATKEKLSAIFGVINETGJHOKRAMPVGVSKANDQILNCLIVNFE IDVSTGVNGLAVRLTPFRKRLHKHNLISRSCLFEQGREIEANVKLTPWPSFIR GFYQWSIGNDISTADHDNNKTSHTFTKTSAPFGSAVANNP" 2359..3077

similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="polymorphic membrane protein G family"  
/protein\_id="AAF38166.1"  
/db\_xref="GI:7189237"  
/translation="MKYSLPMLTSSALVFSLHPMLMAANTDLSNDYENGSSGAAT  
TAKETSDASGTYTLTSLVSNISATIPADKSCITNGALSPFGASHLYQRIAL  
THDGAINNUNTLSEFGSSSLIDSAAPVGTSGGKALICVNTGEGATVFDNMAST  
LOKTSERDGAIVASYSIDIALKTAAALDDORTSKNGKALCSTANTTYOQSGVTF  
SNRTATDGGGITYSKEDOSTLDANTGVYTFSTNKTGAMSDNLLALGTQVLFQ  
ENKTGSAQAQNPFGCCGALICCYLATATDKGLAISQNEISFTNNTTGAAGALIA  
TKCTLDGNTTLFDNDNTATAGCGAIVETEDFSLKSGSTGVTFSTNKTAKGALYSK  
GNSLGTGNTLSEFGSSSLIDSAAPVGTSGGKALICVNTGEGATVFDNMAST  
SLSGNTATVSGAIVATKALHGNLTLEFDGATATAGALIVETEDFSLKSGSTGV  
FSTNKTAKGALHGNLTLEFDGATATAGALIVETEDFSLKSGSTGVFSTNKTAKG  
IATKSLTLENESLFINNTAKRSGGIVAPRCVYSGSESLIFDQNETSETSGALYSK  
NLSITANPVSFTNNSGCGAIVTADSGELSLAIDDTITFSGNRATEGSTPNSIH  
LGAGAKITKLAAPAGHTIYEDPIYMEAPASGTLIELVINPVKAIYPPDPKNGPI  
ASVPVYAPANPNTGIVFSSGKLPISODASIPANTTILNKKILAGANVYKLGAT  
LOVYFPTQDPSTVFNDAAGTLETTNTNDGSLDKNLVSLDLGKRMITLIVANS  
TSGGLKISGDLKFNHNGSFYDNPGLKANLNPFLDLSSTGTVLDDENPSPSMAA  
PDVYQGSWTLVPRKAGAGKTYLVAMQALCTPELPELALIVPSLNNAYNHSIQ  
OBIATAMSDAPSHPEITWIGIGNAHODKQENACFRLISRTIVGSMHTYQETFA  
VAFSQLPFSKSDYVSDIRSOVYAGSLAQSRYVPLHSLRHLVSLVPLPETP  
LVHGOVSYGRHNMHTTKLANNTGKSDMSHSPAVHVGSLPYDLNRYLSTSPY  
VLYGVYVNOKGFQEVADPRIFEDASHVNVNIPMLGFKHESAPSPSALLITLYAV  
DAYRDPHCLTSLTNGTSMSTPATNLSHQAPFAEASHGLKLHGLDCEFASSGELRSS  
SRSYANCGTRSPF"

BASE COUNT 2726 a 2051 c 2367 g 2882 t  
ORIGIN

Query Match 17.4%; Score 558.2; DB 1; Length 10026;  
Best Local Similarity 52.7%; Pred. No. 3e-107;  
Matches 1432; Conservative 0; Mismatches 1228; Indels 57; Gaps 8;

QY 330 AACGTTACTCCAAAACTTCAGCCACAAATTCCTTAACAGAGATCTTCTTTTA 389  
DB 2757 AAATCTACAGCAGATGCCAATGGAAGCAATGCTCTATCAGAAATGCTATATAA 2698  
QY 390 CGAGCCTGAAAAAGCACCCTTATCTGACAGTGTTTTAAACAACACGACAAATCT 449  
DB 2697 CGATGCTGGAAAGGCGACATTAACAGCTGCTCTTTTACGAAAGTGGGTATCT 2638  
QY 450 TACCTTCTTGGGAGAGCATAGCTTAAGCTTGTGCTTATAGATGCTGGACATCATGC 509  
DB 2637 GACATTACTGGAAGAGGATACATTCATTTCAACACGATGATGCGGTTCCGATGC 2578  
QY 510 AGGTGCTGCTGATCTACAAACAGCAAAATTAAGATCTTACCTTCAGGGTTTCTTACT 569  
DB 2577 AGGAGCTGGGCAAGCACAAGCTGATTAAGCCCTACATTCACAGATTTTCTAACCT 2518  
QY 570 GAGTTTGAATTCCTCCACAGCAACAGGTACTACAGCTGAGGAAACGCTTCCACAC 629  
DB 2517 TTCCTTCACTGACGCTCGGAACTACAGTGTCTCAGAAAAAGTACTTAACTTCTGC 2458  
QY 630 AGGAGCGCTAAATTTAGAAATATTCGTAACCTTGATGTGCGGAATTTTCTACTGC 689  
DB 2457 AGGAGCTTAAATCTTACGATATAGGAAGATTCCTTTAGCCAAAGCTCTCCAAATGA 2398  
QY 690 AG-----ATGGTGAAGCTATCAAGAGAGCTTCTTCTTTAAGTGGACATCTTGG 740  
DB 2397 ACCTATAACAATGGGCGCATCACCCCAAAAACCTCTTCTATTTCTGGGAATACCTTC 2338  
QY 741 AGATGCTCTTTTATAGAACACTTCATCAACAAAGAGGAGGCAATGTCTACTACAC 800  
DB 2337 TTCTATAACCTTCACTAGTAAGAGCGCAAAAATTTAGTGGGCAATGTCTTCTTC 2278  
QY 801 AGGCGCTGCATAGCAATACAGAGTATGTAGATCTCTATCAATACATAGCGCTCAC 860  
DB 2277 GGCTGCAAGTATTTTACGAAACACCGGCGCATTTAGTCTTTATAGATTAATAAGAGAAAC 2218

QY 861 GTCAGAGCGCTATGANGATGMAAGCAGTCGATACTATGCAACAAATTTCTATA 920  
DB 2217 TGGGGCTGGGGCTCTGGGGCTTGAAGCCAGCTCTCGATTTACTCAAAATAGCTCCCTTTT 2158  
QY 921 TTTTGAAGGAATGACGCGAAAATCTACT-----GGCGGTGCATCTGCACACCA 971  
DB 2157 CTCTCTGGAACACTGCAACAGATCTCGAGCAAGGCGGGCCATTTATTTGTAAAA 2098  
QY 972 GGGGAGTGAATCCCGCAACGTATATCTTAACATTAAGACATCATGTTTCTTCAAA 1031  
DB 2097 AACAGAGAGACTCTACTCTTCTTCTTCTGGAATTTAAAGCTTAACTTCCCGCAGAA 2038  
QY 1032 CGTAGGAGAAACAGAGGCTGGCGCCATCCATGCTAAAGACTAGCCCTTCTCTGAGG 1091  
DB 2037 CTCTTACGATACCTCAAGGGGAGCATCTGTGCCATTTGGTCTAGATCTTCCGCTCTGG 1978  
QY 1092 CTTTACAGAGTTTCTACGAATATATCTCATACGAACTCT--AAGGGGGTCTAT 1148  
DB 1977 CCTTACCCTATTTTCAAAATATGATGCGGGAACACAGCTGCAGGCAAGCGCGCTAT 1918  
QY 1149 CAGCATGATGCTCGAGAGAGCTGCTTCTGACAGACAGAAACATTTACTTTGT 1208  
DB 1917 TGCATTTGCCGACTCTGATCTTTAAGTCTCTGCAAAATCAAGGACATCACGTTTCT 1858  
QY 1209 AAGAAATACCTTACCAACACCGGAATACCATCTCTTAAACGTAAATCGATTAACAT 1268  
DB 1857 TGGCATATCTTAACCTCAAC--TCCCGCCACATCGACAGGAAATGCTATACCT 1801  
QY 1269 AGGAATACCGGAATTTACGGAATTTACGGGCTGCTAAATCAATCAATTTTCTTCA 1328  
DB 1800 GGGATGTCAGCAAAAATTAACGAATTAAGGCGAGCCAGGCAAGCAATCTATTTCTA 1741  
QY 1329 TGATCCATCACTTCTGAGAGAAAC--CTCATCAGACGTATTGAATTAATTAACGGCTC 1385  
DB 1740 TGATCCGATGATCTTAACACACACAGAGCTTCAAGCTTGTGACATCAACCAACCGGA 1681  
QY 1386 TGGGAGGCTCTCATCATCAATCAATCAAGAAAGATTTATTTTCTGAGAAACCTTAACAC 1445  
DB 1680 TAGCACTCGCTTATGATTAATCAAGAAAGATTTATTTTCTGAGAAAGCTCTCTGC 1621  
QY 1446 AGATGACCTTAAGTGTGACAAATTAATCTCATACGACACCGCTCTCCCTATC 1505  
DB 1620 AGATGAGGAAAGCTGATTAATCACTCACTATATTAATTAAGCAACATTTGCTCTACG 1561  
QY 1506 CGAGGAGAAATTTATTCCTCAAAAGGAGTCACTTATAGAGACAGCAGCTCTCTCAAA 1565  
DB 1560 CTCTGGAACCTTACGACTCAAGAAAGATGTGAGTATGATGATGTTTCAACAGAC 1501  
QY 1566 GGCGGTTCTCTCTGCGATGAGATTCAGAAACGACATTAATCAATCAAGCTGGAGAT 1625  
DB 1500 TGAAGGCTCTACACTCTCATGCAACAGAAACAAAGCTCAAAAGCAGATTAAGTAACT 1441  
QY 1626 TCAATCAAGAACCTTGAATCAATGATGATCTGATGATGATTAACACACCGCTCAAGCT 1685  
DB 1440 CAGCTTACCAAACTTGTGCTGATCTTCTGCTAGAGGAAATTAAGATGTGCTCAT 1381  
QY 1686 AACAGCAAAAGTGTCTCAAAATTAAGTATCTGGAAGCTCAACCTGATTAAT 1745  
DB 1380 TGAACACAGAGGAGCAACAAACATTAATCTTAACCTCTCTCTGTTTCCAAAGATG 1321  
QY 1746 TGAAGGAACTTTATGAAAGTCATATGTTTCAAGCAGTACAGCTTCTCTATTTAA 1805  
DB 1320 TGAAGGCAATTTTATGAAAGCATACGATTAACCAAGCTTCAAGCAGCTTGTGCTG 1261  
QY 1806 AATCAAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1865  
DB 1260 ATTTAC--TGTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1204  
QY 1866 TGTGAGATCTTAATTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1925  
DB 1203 AGTACAAATCTCAGAAACCTCATTTACGGGATACAGGACATTTGGAAGCAGCTTGGGCA 1144  
QY 1926 GGATACAGCTTACAAATTAACAAAGAGGCGCAAGCTTGGACCAAAACAGATTTGTTCC 1985

Db 10069 ATGGAACGACACTATACCTCATGTGCAGAAATCCGAAAGTCTCTGGAGCAGACGATGTATA 10128  
 QY 2521 GCTGTGGTATCGGCTTAGACCTTCCTTTTCTTTCCAAACCCACATCCTTTTCAAG 2580  
 Db 10129 GCTGTGGTATCGGCTTAGACCTTCCTTTTCTTTCCAAACCCACATCCTTTTCAAG 10188  
 QY 2581 ACCTTCATCCACAGATGAAGTGAATGTTATGATACAAATAATAGTCTTCGAA 2640  
 Db 10189 ACCTTCATCCACAGATGAAGTGAATGTTATGATACAAATAATAGTCTTCGAA 10248  
 QY 2641 AGCTTAGTAGTGGCCGTGTTTATGATGGAAGGCTTAAACCTTCGATTCCTGTG 2700  
 Db 10249 AGCTTAGTAGTGGCCGTGTTTATGATGGAAGGCTTAAACCTTCGATTCCTGTG 10308  
 QY 2701 GGTGCGAAATCGGACGGGGGATATCGAATCTCTACACCTTGTGATTCAGGATTC 2760  
 Db 10309 GGTGCGAAATCGGACGGGGGATATCGAATCTCTACACCTTGTGATTCAGGATTC 10368  
 QY 2761 TTTGTTCCGATGCTCTATCTGTAACAAATCCCAATCTACAGGACTCTTGTGATGACCCA 2820  
 Db 10369 TTTGTTCCGATGCTCTATCTGTAACAAATCCCAATCTACAGGACTCTTGTGATGACCCA 10428  
 QY 2821 GACTCTTGGAATAATTCGGGTGGCAATCTTCAAGACAGGCAATTTTACTGAGGGTAGC 2880  
 Db 10429 GACTCTTGGAATAATTCGGGTGGCAATCTTCAAGACAGGCAATTTTACTGAGGGTAGC 10488  
 QY 2881 AACCACTACGCTCTACAACTCCAAATTTGAGAGCTCTTGGACATCTACGCTATGGAACCTCGT 2940  
 Db 10489 AACCACTACGCTCTACAACTCCAAATTTGAGAGCTCTTGGACATCTACGCTATGGAACCTCGT 10548  
 QY 2941 GGATCTTCAAGAACTACAAATGTGATGTTGGTACCAACTCCGATCTAGATTGCTTAA 3000  
 Db 10549 GGATCTTCAAGAACTACAAATGTGATGTTGGTACCAACTCCGATCTAGATTGCTTAA 10608  
 QY 3001 ACTCCCTAGTCTTCTAGGAGGATTTCTCATATCTTTAGGGAATAATTTGCTATAGGGAA 3060  
 Db 10609 ACTCCCTAGTCTTCTAGGAGGATTTCTCATATCTTTAGGGAATAATTTGCTATAGGGAA 10668  
 QY 3061 TGTCTTCTCTGCAACGTGTAAAAATTAACATTTGCTCCCTCTTCAAAAAAGATTCTTTTA 3120  
 Db 10669 TGTCTTCTCTGCAACGTGTAAAAATTAACATTTGCTCCCTCTTCAAAAAAGATTCTTTTA 10728  
 QY 3121 ATAAATCTAGTATTAATTTATTTTAA 3149  
 Db 10729 ATAAATCTAGTATTAATTTATTTTAA 10757

RESULT 7  
 AE002193/c  
 LOCUS 10026 bp DNA BCT 30-MAY-2000  
 DEFINITION Chlamydia pneumoniae AR39, section 26 of 94 of the complete genome.  
 ACCESSION AE002193 AE002161  
 VERSION AE002193.2 GI:8163406  
 KEYWORDS  
 SOURCE Chlamydia pneumoniae AR39.  
 ORGANISM Chlamydia pneumoniae AR39.  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE 1 (bases 1 to 10026)  
 AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
 TITLE Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39  
 JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
 MEDLINE 20150255  
 PUBMED 10684935  
 REFERENCE 2 (bases 1 to 10026)  
 AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,

COMMENT  
 FEATURES  
 SOURCE  
 TITLE  
 JOURNAL  
 Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
 Direct Submission  
 Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
 On Jun 1, 2000 this sequence version replaced gi:7189234.  
 Location/Qualifiers  
 1..10026

/organism="Chlamydia pneumoniae AR39"  
 /strain="AR39"  
 /db\_xref="taxon:115711"  
 /note="synonym: Chlamydia pneumoniae AR39"  
 /complement(105..2897)  
 /gene="CP0307"  
 /complement(105..2897)  
 /gene="CP0307"  
 /note="similar to GP:4376729; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="polymorphic membrane protein G family"  
 /protein\_id="AAE38164.1"  
 /db\_xref="GI:7189235"  
 /translation="MKIPHLKLLISSTLVPTILSIATYGADASLSPDSEFGAGST  
 FTRPKSTADANGTVYLSGANVINDAGCTALTCCEFTTGDLTFTEKGSFSENV  
 AGSNAGAASTTADKALTFPTGFSNLSFIAPAGTVAAGGSLASGALNTDGTILF  
 SONVSNANNGCAITPAKTLSISGNTSITPTNSAKKIGCAIYSSAAISCTNGOL  
 VFMNKGTEGCGALGFPAASSITTONSSLEFSGNATDAKGAATTEKTEGEPITLI  
 SGNKSLTEFAENSSVTQGAICAHGLDLSAAGPLFSNNRGNTAAAGGGAIALADSG  
 LSLSANOGDITFLGNTLSTAPSTNATILGSSAKITMLRAAGOSIYEPYIASN  
 TTGASDVLATINODPSNPLDYSGITVPSGKLSADEKADNFTSILKOPALASGL  
 ALKGVNELDVNGFTQTEGSLTLMOPGKLRKADREALSLKLVYDLSALEGKNSIET  
 AGANKTITLSPLVPOSSGNPFYESTHINAPQPIVAPTAASQIYIDALLTSY  
 QTEPHIGIQHWEATWADTSTAKSGMTVITGYNPPRRASVYDSDIASFTDIR  
 TLDOIMTSQANSIYQOGLWASGTANEFHDKSGTQNAFRKSGYIVGGSADPSEN  
 IFSAVFCOLFGRKDLFIVENTSHNYLASLYLQHRVFLGLPMPSEFSITDMKIDPL  
 ILNVALSLSYTKNDMDRTYRYPEAGSGMTNNSGALDELGSLALYLRKAPFQGYR  
 FLKRVAYSRQONPKESGAERAPDGDIVNCIPVGRLEKISEDKNPFETSLAVT  
 GDYVRKAPRSRTSLMWSGASWTSLCKRLAQAFLASGSHLTLSLPHVELSGEAYELR  
 GSAHLYNDCGLRISF"  
 /complement(2941..5751)  
 /gene="CP0308"  
 /complement(2941..5751)  
 /gene="CP0308"  
 /note="similar to GP:4376728; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="polymorphic membrane protein G family"  
 /protein\_id="AAE38165.1"  
 /db\_xref="GI:7189236"  
 /translation="MKSSVSWLFPSSIPLEFSLSIYAABETLDSNNNSYSGNGTFT  
 VFSITDAAAGTYSILSDVSFQNAAGATIPLAGCELEAGDGLTFQGNOLALFPALN  
 AGSSAGVAVASTSADKNLFLPNDPSRLSIICPSILSPICGACALSGVNLGNSOI  
 IFQNFESDNGGAVYFNKFLISGTSQAFSPKRAKNSLNLAVNLSTIGALSTLTL  
 VFSFSONAKSGGALYSTNCSITDNQVIFPDGNSAFEAAGAGGAAICCTTTNSPVL  
 TGNKNLSFTNNATLTYGASLGKYSISAGPILFQSNISGSAAGCGGCAINIASG  
 ELALSATSGDITFRNNQVITNGSTSRNAIITPAKTSITRAATGSGIYIYDITPBG  
 TAASTDPLNLNLADANSEIYGAIYFSGIKLSPTEKAIAANVTSTIROPAVLARGDL  
 VLARDGVAVFPKDLTQSGRSLMDGCTTSLAKSANSLNLAVNLSTIGALSTLTL  
 EADKNLSISGTLIDTEGSEFNHKLKASQYPLLELTGATGANGATTTGALSTLTL  
 OPEPTHYGGCNGNLSMANATSSKISINMTPRGYIPSPERKSNLPLNSLWAGFTIDR  
 SINOLIEFKSGSEPERELNLSIANFYDSDMPTRHPIHISGVALGTTATTPED  
 QLTFAFQDLFARDRNHTGKNHCDTIGASLYFHHTBLGFIANFLWKATRAVAVSE  
 LSQIILPISFPAKPSYLTNDNMKTYYTDNLSIIRGSMDNFCADLGSLEFVLSVPL  
 LKEVEPFGQVYIYAHQODEYERYAEGRAPNKSELINVEIPDGVTEFRSDSKSEGYD  
 LTLTALIDAVARNPKCOTSLIASDMMMAVGTNLARQGSFVRANRHQVAPHMEIRGO  
 FAREVSSSRNVTNLSKRCF"  
 /complement(6013..9843)  
 /gene="CP0309"  
 /complement(6013..9843)  
 /gene="CP0309"  
 /note="similar to GP:4376727; identified by sequence



QY	301	GAGGCTTTATGGAATATCGAATTCAGAGAACGTTTACTCCAAAAAAGCTTGACCCACACA	360
Db	7909	GATAGCTTTAATGGAAATATCGATGATGAGAACGTTTACTCCAAAAAAGCTTGACCCACACA	7968
QY	361	TATCTCTTAACAGAGAGATGTCTTCTTTTACGAGCCGGAAGAGCACTCCCTATCTGAC	420
Db	7969	TATCTCTTAACAGAGAGATGTCTTCTTTTACGAGCCGGAAGAGCACTCCCTATCTGAC	8028
QY	421	AGTGTGTTTTAAGCAACCGACGACAATCTTACCTCTGSGGAACGGCTATAGCTTAAAG	480
Db	8029	AGTGTGTTTTAAGCAACCGACGACAATCTTACCTCTGSGGAACGGCTATAGCTTAAAG	8088
QY	481	TTTGCGCTTTATAGTGTGTGGCACTCAATGCAAGGGTGTGCGATCTCAACACCAATATAG	540
Db	8089	TTTGCGCTTTATAGTGTGTGGCACTCAATGCAAGGGTGTGCGATCTCAACACCAATATAG	8148
QY	541	AATCTAACCTCTCAGAGGTTTCTTACAGATGATTTGATTTCTCTCTAGCAACAGGTT	600
Db	8149	AATCTAACCTCTCAGAGGTTTCTTACAGATGATTTGATTTCTCTCTAGCAACAGGTT	8208
QY	601	ACTACAGGTCAGGACGCTTTCCTCAGCAGAGGCGTAATTTAGAAAATATTCGTAAA	660
Db	8209	ACTACAGGTCAGGACGCTTTCCTCAGCAGAGGCGTAATTTAGAAAATATTCGTAAA	8268
QY	661	CTTGATAGTGTGGGAATTTTCTACTCGAGATGGTGGAGCTATCAAGGAGCGCTTTC	720
Db	8269	CTTGATAGTGTGGGAATTTTCTACTCGAGATGGTGGAGCTATCAAGGAGCGCTTTC	8328
QY	721	CTTTTAAGTGGCACTTGTGAGATGCGCTTTTATGAAACACTCTTCATCAACAAAGGGA	780
Db	8329	CTTTTAAGTGGCACTTGTGAGATGCGCTTTTATGAAACACTCTTCATCAACAAAGGGA	8388
QY	781	GGAGCAATTGTCTACTACAGCAGGCGCTCGATAGCAAAATACACAGGTTATTTAGATTTC	840
Db	8389	GGAGCAATTGTCTACTACAGCAGGCGCGGATAGCAAAATACACAGGTTATTTAGATTTC	8448
QY	841	CTATCTTAACATAGGCTTACGTACGAGAGCGCTATGTGATGAAAGCACTGCTACTA	900
Db	8449	CTATCTTAACATAGGCTTACGTACGAGAGCGCTATGTGATGAAAGCACTGCTGATCTA	8508
QY	901	TCGAAACAACAATTTCTATATTTTGAAGGGAATGCAGCGAAACATCTGGCGGTCCGATC	960
Db	8509	TCGAAACAACAATTTCTATATTTTGAAGGGAATGCAGCGAAACATCTGGCGGTCCGATC	8568
QY	961	TGCAACACCAAGGCGAGTGGATCTCTGCACTGATATCTCTACAAATTAAGACTTGTGATC	1020
Db	8569	TGCAACACCAAGGCGAGTGGATCTCTGCACTGATATCTCTACAAATTAAGACTTGTGATC	8628
QY	1021	TTTGCTTCAAACTGACGACAACAGCGGTGGCGCCATCCATCTPAAAAAGTACGCCCTT	1080
Db	8629	TTTGCTTCAAACTGACGACAACAGCGGTGGCGCCATCCATCTPAAAAAGTACGCCCTT	8688
QY	1081	TTCCCTGTGAGGCTTTACAGATTTTTCACGAATTAATGCTCATGCAACTCTTAAGGGG	1140
Db	8689	TTCCCTGTGAGGCTTTACAGATTTTTCACGAATTAATGCTCATGCAACTCTTAAGGGG	8748
QY	1141	GGGTCTATACAGCATCGATGGCTCAGAGAGACTAGCTTTTCGACAGACAGAGAAACTT	1200
Db	8749	GGGTCTATACAGCATCGATGGCTCAGAGAGACTAGCTTTTCGACAGACAGAGAAACTT	8808
QY	1201	ACGTTTGTAGAAAATACCTTTACAAACACCGGAAGTACCGATACTCTPAACGTAATGCG	1260
Db	8809	ACGTTTGTAGAAAATACCTTTACAAACACCGGAAGTACCGATACTCTPAACGTAATGCG	8868
QY	1261	ATCAACATTAAGGAAGTAAACGGGAATTCACGGGAATTAACGGGCTGTCAAAAATCTATCAATT	1320
Db	8869	ATCAACATTAAGGAAGTAAACGGGAATTCACGGGAATTAACGGGCTGTCAAAAATCTATCAATT	8928
QY	1321	TTTCTTATATATCCCATCACTTTCAGAGAGAACCTCATCAGAGCTATTAAGATTAATATAC	1380
Db	8929	TTTCTTATATATCCCATCACTTTCAGAGAGAACCTCATCAGAGCTATTAAGATTAATATAC	8988
QY	1381	GGCTCTGGGGAGCTCTCAATTCATATCAAGAGACGATTTCTATTTTCTGGAGAAACCTTA	1440

Db	8989	GGCTGTGGGGAGCTCTCAATTCATATCAAGAACGATCTTAATTTTTCTGGAGAAAACCTTA	9048
Qy	1441	ACAGCAGATGAACTTAAAGTTGCTGCACAATTTAAATCTTCATTACAGCAGCCAGTCTCC	1500
Db	9049	ACAGCAGATGAACTTAAAGTTGCTGCACAATTTAAATCTTCATTACAGCAGCCAGTCTCC	9108
Qy	1501	CTATCCGGAGGAAATTTTGTCTACAAAAGGAGTCACTTTAGAGACACAGATTTCTCT	1560
Db	9109	CTATCCGGAGGAAATTTTGTCTACAAAAGGAGTCACTTTAGAGACACAGATTTCTCT	9168
Qy	1561	CAAGAGGCCGGTCTCTCTCGGCATGTGATTCAGAACGACATTAATCAACTACAGCTGGG	1620
Db	9169	CAAGAGGCCGGTCTCTCTCGGCATGTGATTCAGAACGACATTAATCAACTACAGCTGGG	9228
Qy	1621	AGTATTACATCAGCAACCTAGGAATCAATGTGACTCTTAGTCTTTAAGCAGCCGTC	1680
Db	9229	AGTATTACATCAGCAACCTAGGAATCAATGTGACTCTTAGTCTTTAAGCAGCCGTC	9288
Qy	1681	AGCCTAACAGCAAAAGTGCTTCAAAATAAAGTGCATATCTGGGAAAGCTCAACCTGATT	1740
Db	9289	AGCCTAACAGCAAAAGTGCTTCAAAATAAAGTGCATATCTGGGAAAGCTCAACCTGATT	9348
Qy	1741	GATATTGAAAGGAACATTTATGAAAAGTCAATPMTGTACCCATGAGCCAGCTTCTCTCTA	1800
Db	9349	GATATTGAAAGGAACATTTATGAAAAGTCAATPMTGTACCCATGAGCCAGCTTCTCTCTA	9408
Qy	1801	TTAAAAATCAGGGTTGATGCTGATGTTGTGATCTAACTGTGATCAGCAGCCCTTATCCT	1860
Db	9409	TTAAAAATCAGGGTTGATGCTGATGTTGTGATCTAACTGTGATCAGCAGCCCTTATCCT	9468
Qy	1861	GTTCTGCTGAGAGATCTTAATTCAGAAATAGSGATTCCAAAGACAAATGGAATTTGG	1920
Db	9469	GTTCTGCTGAGAGATCTTAATTCAGAAATAGSGATTCCAAAGACAAATGGAATTTGG	9528
Qy	1921	ACTAGGGATPACAGCTACAAATACAAAAGAGCCAGCGCACTTGAGCCAAAAGAGATT	1980
Db	9529	ACTAGGGATPACAGCTACAAATACAAAAGAGCCAGCGCACTTGAGCCAAAAGAGATT	9588
Qy	1981	GTTCCACACCCCGAAGAAATCTGCTTACTATGCAATACCTCATGSGAGCTTTTACT	2040
Db	9589	GTTCCACACCCCGAAGAAATCTGCTTACTATGCAATACCTCATGSGAGCTTTTACT	9648
Qy	2041	GACATTTGCTCTCTGCAACAGCTGTGTAGAGATCGGGCACTGTTATGGAACACAAAACAA	2100
Db	9649	GACATTTGCTCTCTGCAACAGCTGTGTAGAGATCGGGCACTGTTATGGAACACAAAACAA	9708
Qy	2101	GGTTTCTGGGTTTCTCCATGACCAACTTCTGCATAGAGACTGGAGATGAAAATCGCAAA	2160
Db	9709	GGTTTCTGGGTTTCTCCATGACCAACTTCTGCATAGAGACTGGAGATGAAAATCGCAAA	9768
Qy	2161	GGCTTCGCGTATACCTGTGAGAGGCTCAATCGGTGGAATGTCACACTCTTAAGAC	2220
Db	9769	GGCTTCGCGTATACCTGTGAGAGGCTCAATCGGTGGAATGTCACACTCTTAAGAC	9828
Qy	2221	GACCTATTACCTTTCGCTTCTGCGCATCTCTTGTGTAGACAAAGATTGTTTATGCT	2280
Db	9829	GACCTATTACCTTTCGCTTCTGCGCATCTCTTGTGTAGACAAAGATTGTTTATGCT	9888
Qy	2281	CACAACAACCTAGAACCTACGGTGAACCTTTATTTCAAGCACTCTCATACCTTACAA	2340
Db	9889	CACAACAACCTAGAACCTACGGTGAACCTTTATTTCAAGCACTCTCATACCTTACAA	9948
Qy	2341	CCCCAAACCTATTGAGATTAGGAAGACCAAAAGTTTCTGAAATAGCTATGAAAATTC	2400
Db	9949	CCCCAAACCTATTGAGATTAGGAAGACCAAAAGTTTCTGAAATAGCTATGAAAATTC	10008
Qy	2401	CCTAGGAAATTTCCCTAGCCTTGGAATCTCCAAAGTTTCTGACACCATTCAGACAACCGT	2460
Db	10009	CCTAGGAAATTTCCCTAGCCTTGGAATCTCCAAAGTTTCTGACACCATTCAGACAACCGT	10068
Qy	2461	ATGGAAGCAGCATTAATCTCAATTGCGAGAAATCCGAAGTTTGTGGAGCAACGATGTATA	2520



QY 3121 ATATTTCTAGTATATTTTAAAAACAGTTAAATTAATAGACAATATCTA 3180  
|||||  
Db 210909 ATATTTCTAGTATATTTTAAAAACAGTTAAATTAATAGACAATATCTA 210968  
QY 3181 TTTCTATTGACTCTTTT 3200  
|||||  
Db 210969 TTTCTATTGACTCTTTT 210988

RESULT 6  
AE001628 10757 bp DNA BCT 01-DEC-2000  
LOCUS Chlamydia pneumoniae section 44 of 103 of the complete genome.  
DEFINITION AE001628 AE001363  
ACCESSION AE001628.1 GI:4376730  
VERSION  
KEYWORDS  
SOURCE Chlamydia pneumoniae CWL029.  
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
REFERENCE 1 (bases 1 to 10757)  
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,  
Olinger, L., Grimwood, J., Davis, R.W., and Stephens, R.S.  
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
MEDLINE 99206606  
PubMed 10192388

REFERENCE 2 (bases 1 to 10757)  
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,  
Grimwood, J., Davis, R.W., and Stephens, R.S.  
TITLE Direct Submission  
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University  
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA

FEATURES  
source  
1..10757  
/organism="Chlamydia pneumoniae CWL029"  
/strain="CWL029"  
/db\_xref="taxon:115713"  
106..2892  
/gene="pmp\_9"  
/note="Cpn0447"  
106..2892  
/gene="pmp\_9"  
/codon\_start=1  
/product="Polymorphic Outer Membrane Protein G/I Family"  
/protein\_id="AADI8591.1"  
/db\_xref="GI:4376731"  
/translation="MKSSLHFWLISLSPALPLSNFSAFAVEINLGPNSGPGCT  
YPPAQTNADCTIYNLTGDSITNAGSPATLASCFTKGTGNSLFGHGQYOLFONI  
DAGNCTPNTANKLISFGFSYLSLQITNATGCTGAKSGAGSIOSSCYFGO  
NFSNNGALOGSSISLSPMLTFAKNATQCGALYTGCGITNTNLSFSENT  
AANNGAITEASSPISNNKALISFINNSTATISATGALYCGSTSPKPYLSDNGE  
LNFIGNATISGALITDNLVLSGGPTLFKNNSAIDTAAPLGAIALDGLSLSLA  
LEGDITFEGNTVAKGASSQTTNRSINGNNAKIVQLRASQNGTIVPYDTISIT  
ALSLDALNNGDLGNPAVQGTIVSGEKLSEAEAEADNLSSTIOQPLDTAGGSS  
LKSQVTLVAKSFGSPGSLTLDAGTLEAGTITNNLVNDSLEKFKKATLAKTO  
ASQVTLGSLSLVPSGNVYDVSNMNNQVSCITLTPDDANHIHTPLADLEKN  
PIHMGYQCNMAISNOEDPATRKKAALITTKTCYNNPERRGTIVANTIMGSPYDVS  
IOQLVAIKVROSQETRGINCEGISNFEHSDSTKINGFHHISGALVGAATLLASDL  
ITAAFCQLEGRDHRHINKRASAYAAASHLQHLATLSSPLRLYLGSESQPLVDL  
AOISYISKNTKRTTYTQAPKGESSWNGCALSLAHPHALSHEGLFAHYFPPIK  
VVASIHDSFKERTTILVRSFSDGLINVSPIGITFERFSRNEASYLEAVIYVAD  
VYRKPNDCITALLINNTSWKTGNTLSRQAGIGRAGIIFYAFSPNLEVTSNLSMEIRGS  
SSSYAADLGGKRF"

gene  
complement(3233..4348)  
/gene="yxjG\_2"  
/note="Cpn0448"  
complement(3233..4348)  
/gene="yxjG\_2"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical Protein"

CDS  
106..2892  
/gene="pmp\_9"  
/note="Cpn0447"  
106..2892  
/gene="pmp\_9"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical Protein"

/protein\_id="AADI8592.1"  
/db\_xref="GI:4376732"  
/translation="MEMSPFOPEQCHFDVYGSFLRPESLTPARSDFEGRIVYROM  
RYVEDAATRNLIKQTEACILFFTPDSEFRYSNDPDPFMGCFHVDRRSRNSPEQYV  
LKDKISVSKHPTEHFEVTEKGNNAKQITPSSQFHEHTAPNLRKTRKPTPT  
NOELIDDIIVYRYOVIQDLYAAGCRLNLODDCAWCLLDIIRAPSWGVSHRLOETL  
EQLFVHNLVMDKRPEDLEFVSLHVCGRDQAEFFSRRAVDSIEEPLFAATDVDSHY  
MALDXYSGAEPLAVSGEKHYCLGISNSHSCIEDRRAVVSRIYEAAVSIPLERLS  
LSPOGFASCEGDHMRTEEMOKKIAFVAEIKKEIWC"

gene  
complement(4865..6769)  
/gene="pmp\_10.1"  
/note="Cpn0449"  
complement(4865..7650)  
/note="Polymorphic Outer Membrane Protein G Family;  
contains frame shift"  
complement(6715..7650)  
/gene="pmp\_10.2"  
/note="Cpn0450"  
7813..10599  
/gene="pmp\_11"  
/note="Cpn0451"  
7813..10599  
/gene="pmp\_11"  
/codon\_start=1  
/transl\_table=11  
/product="Polymorphic Outer Membrane Protein G Family"  
/protein\_id="AADI8593.1"  
/db\_xref="GI:4376733"

CDS  
7813..10599  
/gene="pmp\_11"  
/codon\_start=1  
/transl\_table=11  
/product="Polymorphic Outer Membrane Protein G Family"  
/protein\_id="AADI8593.1"  
/db\_xref="GI:4376733"

BASE COUNT 3228 a 2459 c 2065 g 3005 t  
ORIGIN

Query Match 98.4%; Score 3149; DB 1; Length 10757;  
Best local Similarity 100.0%; Pred. No. 0;  
Matches 3149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATGTCGAAGAGAGACTACACAGGAAATTCGATTCATTAACCCACTTATTATTA 60  
|||||  
Db 7609 CAATGTCGAAGAGAGACTACACAGGAAATTCGATTCATTAACCCACTTATTATTA 7668

QY 61 AATTCCTACTGGCATATAAATAAGAAACACACAGAGTCAAGATTAATAATCTTGAC 120  
|||||  
Db 7669 AATTCCTACTGGCATATAAATAAGAAACACACAGAGTCAAGATTAATAATCTTGAC 7728

QY 121 AGCTGTTTGTCACTCTTAACTGATTTACTTATTTGTTTCTATATTGATGCGAATAGT 180  
|||||  
Db 7729 AGCTGTTTGTCACTCTTAACTGATTTACTTATTTGTTTCTATATTGATGCGAATAGT 7788

QY 181 TCTCTAAAAAACAAGACATTAACATGAAGACTTCCTGGATTGTTTCTCTCC 240  
|||||  
Db 7789 TCTCTAAAAAACAAGACATTAACATGAAGACTTCCTGGATTGTTTCTCTCC 7848

QY 241 GGTGTAGCTTCTCATGTACCTACAGTCACTAGCTAACGAGAACTTTATACCTGAT 300  
|||||  
Db 7849 GGTGTAGCTTCTCATGTACCTACAGTCACTAGCTAACGAGAACTTTATACCTGAT 7908



```
/db_xref="GI:8978646"
/translation="MHLEENOGMEALLRKVYHOEVPATILHGFLLPVLQDAEOLA
SETLSSSSPGESEKHSOKIHDPIDYOPPEKGRSLSDIPRGKIKOYISPPRANKYI
YIIHEDRMTLAIISAFKVEEPKHAVIDITTAQKQRIKPTIISLSITFERGK
ILCSKETSFLEFRYAOCEIPVTEVSDIKESSEFDQVLEKQREMEVLELEYRDY
TILMLGKASALANYPEHVEKITIOLPLPLPDLKVLIVESASRLNSSSAASLEWVAIO
LVSLOYKEKELVSVSPGODLSN"
complement(5246..5866)
/gene="tdk"
complement(5246..5866)
/gene="tdk"
complement(5246..5866)
/gene="tdk"
/translation="thymidylate kinase"
/product_id="BAA98483.1"
/codon_start=1
/transl_table=11
/protein_id="BAA98483.1"
/db_xref="GI:8978647"
/translation="MFLVIEGGEGSGKSLAKALGDQVADNRKVLITREPGLIGE
RLMDLLEPPELHLSRCELEFLGSAHQHIOVYIIPALDGYIVICERPHDSTIYO
GIAEGGADVFADICSKVGPPELPNVEVLIDIPADIGORRHQCFEPKPKPLS
YHNRIRGFLSLASADPSRYLVADRESLASLIDKVLHTQLGLCT"
complement(5871..8375)
/gene="gyrA.1"
complement(5871..8375)
/gene="gyrA.1"
complement(5871..8375)
/gene="gyrA.1"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit A"
/protein_id="BAA98484.1"
/db_xref="GI:8978648"
/translation="MFNKDEIIVPKNLEEMKESYLSNSVLIISRALPIDRLKPS
ORVLVAMKOLSLSPGAKHRKCAKICGDTSGDHDVLEPVMAMQAMRYEL
VDGONGSISDGPAPAMRYTEADRLVLEMDKDVDPVLPVNDKEHPVEP
SKEPNLICNGSSGIAVGMATNPNNHLEIETLILILAPROSVDEILOVMGPEP
TGGIIGSEBISRTYTGKRTIKVRAHLHEENEDKRESITITTEPNYKRLRGL
IHALVETKLAGISDVDESDKDGIRVLEIKGESSEIINLRYFTVQVFLGAM
LADKNI.PRTMSIHRMSAMIRHREKVIIRRTYELNKATRAHVEGYLAKALSDA
LVKTIRESGNKEHAKERIIIESFGTEPOLAILERLYOLGAEKIOKEVELLAK
IAYKQVLSDEGLKDIIRNELQDLKHKHVARRTIEFPADDIRIEDIITNESYI
TISGDYVYKMPYKVFKEORRGHGTGFMKAGFKAVASAFTRDYLIPTNGO
CYLKYMWOLPEGRBAKRPINPFIKIGRGEELALINKNEDMAGFLFIAKRYV
KRYSLDAFSPNPKRGTRALEIDEGDELIACHIVSBEKYMLETILGMAVRFHEKVR
PMGRVAGVAGVSLKNEEDVVSQIVTEMSVLIVCDGFGKSLVEDREFNNGAV
GVSLILINERNVGLAIPYTHDSILMSQOQARINQDVRVAGKRSIQGVRVLHL
KESDALVSMKESINENDEVLSSSEECSDTVSLR"
complement(8391..10808)
/gene="gyrB.1"
complement(8391..10808)
complement(8391..10808)
/gene="gyrB.1"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="BAA98485.1"
/db_xref="GI:8978649"
/translation="MDPKERNYDASAITYLEGLQAVREPRGYIGDTIGTLHLHYE
LVNDSIDEAMAGYCSRIYVLEIDGIVIVDENGRPIEVEHERESKOGVEYVALEY
LYVLHAGKEFKDSYKSGSLGVLGVCVVALSEKRGIAVYFKKCYOMFSGSIPY
PLQYVSVDRGTEIYVPPDKIESTCTPRISILMKRLRDLINGLITRIGTNDSDY
SPDKYFPYEGGIGQSFVSYLNOKRESLFSPIYTCITRGDDDEIFEPALQNSDYS
ELVYSTANNIPTRGGTHLGEFTALFRVINYIKAHNKLKNNKLALTEGDIETGLA
VISVKNPOPEGOTKOKLNSDVSVAOVVEALTIFFEENPOLARMLVDPVFAA
QAREAKKARELTLRKSAIDSAIRPGKIDLCDEKDEKEMYVEVEDSGAGSKOCD
RRROALIPTRGKILNVEKARLOKIFONOEIGTIIALGGIGADNNTLSLRRII
MTDADVDSHIRTILLTFYRHM TALINENCYVIAQPLKYKKKDFRIISEKEDM
STILMGTNNESSILFKSTERELRGEALSEFINVIDVESTINLEKALIPESFLEM
KEBIGPLVYLAATGMOGRYLYSDEEKREALADETHFKIILELYKVAVFIDIQO
LKEYGDISSYLLPQKNEIYVGNEDSPSCNYSCTYLEEYINYLKNGKRGIEQRYG
LGMANDQJLMDTMINPEORTLLHVLKDAVEADHIFTMLGEEVPPRREIESHALSI
RINNDI"
complement(10812..11159)
/gene="CPJ0276"
complement(10812..11159)
/gene="CPJ0276"
/codon_start=1
```

```
/transl_table=11
/product="CPJ91 hypothetical protein"
/protein_id="BAA98486.1"
/db_xref="GI:8978650"
/translation="MFLKRRKRGSOVONKRTASPIKHAHYLHNTLQELQIMARP
HDAIDAMNOVFRDKYKGMQSOAIFRHHILLVKVNSLYALLKOTPDNDLMSLYOVA
SHVOIREIOPFLG"
Query Match 100.0% Score 320.0; DB 2; Length 300650;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATGCGAAGAGAGACTAACACAGAAATTCGATTCATTAACCCACTTTATTTA 60
Db 207789 CAATGCGAAGAGAGACTAACACAGAAATTCGATTCATTAACCCACTTTATTTA 207848
QY 61 AATTCCTACTTGGCTCATATAAATAAGAAAATCAGAGAGTCAAGATTAATTCCTGAC 120
Db 207849 AATTCCTACTTGGCTCATATAAATAAGAAAATCAGAGAGTCAAGATTAATTCCTGAC 207908
QY 121 AGCTGTTTGCATCTTAACTGATTTACTATTGTTTCTTATTTGATGATGATAGT 180
Db 207909 AGCTGTTTGCATCTTAACTGATTTACTATTGTTTCTTATTTGATGATGATAGT 207968
QY 181 TCTCTAAAAAACAACATTAACCATGAGACTTCGATTCCTGGTTTATGTTCTCTCC 240
Db 207969 TCTCTAAAAAACAACATTAACCATGAGACTTCGATTCCTGGTTTATGTTCTCTCC 208028
QY 241 GTGTTACCTTTCATGTCACTACCTACCTACCTACCTACCTACCTACCTACCTACCT 300
Db 208029 GTGTTACCTTTCATGTCACTACCTACCTACCTACCTACCTACCTACCTACCTACCT 208088
QY 301 GATGAGCTTAATGAATATCGATTCAGAAAGCTTCTCTCAAAAACCTGACCCACACACA 360
Db 208089 GATGAGCTTAATGAATATCGATTCAGAAAGCTTCTCTCAAAAACCTGACCCACACACA 208148
QY 361 TATCTCTAACAGAGATGCTCTCTTTTACGAGCTTGAAAGAGCACTCCCTTATCTGAC 420
Db 208149 TATCTCTAACAGAGATGCTCTCTTTTACGAGCTTGAAAGAGCACTCCCTTATCTGAC 208208
QY 421 AGTTGTTTAAAGCAACACGAGACATCTTCTCTTGGGGAACGCTCATAGCTTAAGC 480
Db 208209 AGTTGTTTAAAGCAACACGAGACATCTTCTCTTGGGGAACGCTCATAGCTTAAGC 208268
QY 481 TTTGGCTTTATAGTGTGGTGGGACATCATGAGAGTCTGCTGATCTTACACAGCAATAAG 540
Db 208269 TTTGGCTTTATAGTGTGGTGGGACATCATGAGAGTCTGCTGATCTTACACAGCAATAAG 208328
QY 541 AATCTTACCTTTCAGAGGTTTTCCTTACTAGTATTTGATTCCTCTCTTACACACAGGTT 600
Db 208329 AATCTTACCTTTCAGAGGTTTTCCTTACTAGTATTTGATTCCTCTCTTACACACAGGTT 208388
QY 601 ACTACAGGTCAGGACGCTTTCCTCAGCAGAGGCGTAAATTTAGAAAATTTCTGTA 660
Db 208389 ACTACAGGTCAGGACGCTTTCCTCAGCAGAGGCGTAAATTTAGAAAATTTCTGTA 208448
QY 661 CTTCGTACTTGTGGGATTTTCTTACGCGAGTGTGAGTGTGAGCTATCAAGAGAGGCTCTTC 720
Db 208449 CTTCGTACTTGTGGGATTTTCTTACGCGAGTGTGAGTGTGAGCTATCAAGAGAGGCTCTTC 208508
QY 721 CTTTAACTGGCACTTGTGAGATGCTCTTTTATGTAACACTCTTATCAACAAAGGA 780
Db 208509 CTTTAACTGGCACTTGTGAGATGCTCTTTTATGTAACACTCTTATCAACAAAGGA 208568
QY 781 GGAGCAATTCCTACTACAGAGCGGCTCGCATACCAATAACACAGGTTATGTTAGATTTC 840
Db 208569 GGAGCAATTCCTACTACAGAGCGGCTCGCATACCAATAACACAGGTTATGTTAGATTTC 208628
QY 841 CTATCTAACAATAGCGTCTAGTACAGAGGCGGCTATGATATGATGAAGCAGCGATCTACTA 900
Db 208629 CTATCTAACAATAGCGTCTAGTACAGAGGCGGCTATGATATGATGAAGCAGCGATCTACTA 208688
QY 901 TCGAACAAACAATTTCTATATTTTGAAGGAGATGACAGAAACTACTGCGGTGATC 960
```

Db 14211 AACACTACGCTACACACCTGCAATGCTGAGCTCTTGCGACATGACTAGACCTCCGT 14270  
OY 2941 GGATCTTCAAGACACTACATAGTAGTGTGGTACCAAACTCCGATTCAGATTGCTAAA 3000  
Db 14271 GGATCTTCAAGACACTACATAGTAGTGTGGTACCAAACTCCGATTCAGATTGCTAAA 14330  
OY 3001 ACACCCAGTCTCTAGGAGTTTCTCATCTTTAGGAAATATTTGCTATAGGAA 3060  
Db 14331 ACACCCAGTCTCTAGGAGTTTCTCATCTTTAGGAAATATTTGCTATAGGAA 14390  
OY 3061 TCGTTCCTGCAAACTGTAATAATACATTTGCTCCCTCTTCAAAAAAGATTCTTTA 3120  
Db 14391 TCGTTCCTGCAAACTGTAATAATACATTTGCTCCCTCTTCAAAAAAGATTCTTTA 14450  
OY 3121 ATATTTCTAGTATATTTATTTTAAACAGTTAAATATTAATAGCAATATCTA 3180  
Db 14451 ATATTTCTAGTATATTTATTTTAAACAGTTAAATATTAATAGCAATATCTA 14510  
OY 3181 TTCTATGACTCTCTTTT 3200  
Db 14511 TTCTATGACTCTCTTTT 14530

RESULT 5  
AP002546 300650 bp DNA BCF 08-DEC-2000  
LOCUS Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,  
DEFINITION section 2/4.  
ACCESSION AP002546 AB033782 AB033783 AB033784 AB033785 AB033800 AB033801  
AB033802 AB033803 AB033804 AB033805 AB033806 AB033807 AB033808  
AB033809 AB033810 AB033811 AB033812 AB033813 AB033814 AB033815  
AB036079 AB036080 AB036081 AB036082 AB038348 AB038349 BA000008  
AP002546.2 GI:10176692  
VERSION Chlamydomophila pneumoniae J138 (strain:J138) DNA.  
KEYWORDS Chlamydomophila pneumoniae J138  
SOURCE Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
ORGANISM 1 (sites)  
REFERENCE Shitai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,  
AUTHORS Shiba,T., Ishii,K., Hattori,M., Kohara,S. and Nakazawa,T.  
TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CWI029 from USA  
JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)  
MEDLINE 20330349  
REFERENCE 2 (bases 1 to 300650)  
AUTHORS Shitai,M.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.  
MUTATION Mutunori Shitai, Yamaguchi University School of Medicine,  
Department of Microbiology, 1-1-1 Minamikogushi, Ube, Yamaguchi  
755-8505, Japan (E-mail:mshtai@po.cc.yamaguchi-u.ac.jp,  
Tel:81-836-22-2227, Fax:81-836-22-2415)  
COMMENT On Sep 15, 2000 this sequence version replaced gi:6172290  
gi:6172292 gi:6172294 gi:6172296 gi:6172326 gi:6172328 gi:6172330  
gi:6172332 gi:6172334 gi:6172336 gi:6172338 gi:6172382 gi:6172384  
gi:6172386 gi:6172388 gi:6172390 gi:6172392 gi:6172394 gi:6174666  
gi:6174668 gi:6635174 gi:6635176 gi:6635178 gi:6635180 gi:8547433  
gi:8547438 gi:8978640.  
AB033782-AB033785, AB033800-AB033815. Submitted (25-Oct-1999)  
AB038348-AB038349. Submitted (14-Feb-2000)  
AB036079-AB036082. Submitted (18-Dec-2000).  
FEATURES  
source  
1. 300650  
/organism="Chlamydomophila pneumoniae J138"  
/strain="J138"  
/db\_xref="taxon:138677"  
/note="synonym:Chlamydia pneumoniae (strain J138)"  
144. .935  
/gene="CPj0267"  
144. .935  
/gene="CPj0267"  
/codon\_start=1  
/codon\_start=1

gene  
CDS  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA098477.1"  
/db\_xref="GI:8978641"  
/translation="MSIMSLNKTNALNQPEPAVCLNAMPKYLINQDRKTRACTVTL  
VIATLILITGVIVLVLAMGSPDLSTVSTIIGTSTVLTALFIIIGLKLKSLAMI  
OYKYEORVKKOYEPFSLPKNDYHKLSCPLSDIESPSEASTPSKRLICSG  
VAIVGTLILGAVSVFPCGYLOALCVGACGTLAPUGLGLNTHSLINQIM  
YLILYISLALERNETVKQDRNEINYLTEECNQKREKALLE"  
957. .1343  
/gene="CPj0268"  
957. .1343  
/gene="CPj0268"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA098478.1"  
/db\_xref="GI:8978642"  
/translation="MSOCSSSTJWEMKSFVPMKNPTPLSPIDSEDEFILAYEP  
FVLPTDENAOANPGSTPNVENGIDLPNLGQPNQNNANNPGTSGNPTSLPA  
PERLPETRENSQEEBOSONMEDLG"  
complement(1495. .2472)  
/gene="CPj0269"  
complement(1495. .2472)  
/gene="CPj0269"  
/codon\_start=1  
/transl\_table=11  
/product="dipeptidase"  
/protein\_id="BA098479.1"  
/db\_xref="GI:8978643"  
/translation="MTIDMHCDDLSPHFCKRPVAVRCSPEQLISGVQOYCAFLVP  
HSRGCNDKONSLFSPNOYDPLGLSEEESSOKSLISLISEAVALGD  
DPAICGLILAKIHLITKQGLAYLIGVGNDRPCGGEAPRSLNDDKVLIDIVEL  
GVPIIDLSHCSDKLAEDIDYTRDKLPINLAVIASHNPFVSVDHRNNIDYDAKKEIVR  
KGVIGLNVRSVSGSLDLEKHLVHAENLGLISLVGSDFFANDEENPFNCCSS  
AEAPHVNLQILHRIKSGKAEISLSRAEKLKQYIVBOVNRITDVL"  
complement(2487. .3347)  
/gene="ywlC"  
complement(2487. .3347)  
/gene="ywlC"  
/codon\_start=1  
/transl\_table=11  
/product="SuA5 superfamily-related protein"  
/protein\_id="BA098480.1"  
/db\_xref="GI:8978644"  
/translation="MPDKKAQTFSLPEVMSNHQKIVALPTDVGVSIVLASEA  
BERLALDREPSKAFALVNSIEDIENISGPIPLSKAKLAQDEPGAITLVKRNRP  
RFPKPTLAFRIYDHSVREIYDHCGLTIGTSANLSEFSAITAOEIFADPDHDCIF  
DGPCSHGLESTVVASDPLIYREGIISRSVIENLAGEAKIFHRTSHAFSKHIKIYV  
KNOELVSEFLSGSLDFKGVCEHPKRNFTRLREALKKTPSIYFIVDINTSDPEL  
PPLSPVYIE"  
3656. .4387  
/gene="CPj0271"  
3656. .4387  
/gene="CPj0271"  
/codon\_start=1  
/transl\_table=11  
/product="lysophospholipase esterase"  
/protein\_id="BA098481.1"  
/db\_xref="GI:8978645"  
/translation="WTDYSFRRKIGNIEAIECPGNODPIIILCHGYGLADNLTFP  
PSICFSKLRPRTNIPNGILPNDENFRSRACFPLNVLLOELSTLVANGVNLQEKY  
DELFPVDLETREKALEELILNINRYNIIIGGFSOGAILATHVLVTFSONPYAGLIL  
AGARLFGNGWEGELKCOAOVPFIOSHGVEDELIPYHLCAHNLNLLTLNQGVSFHC  
GHEIPSVFQKQVYVPMWIDPARC"  
complement(4359. .5255)  
/gene="dnaX\_2"  
complement(4359. .5255)  
/gene="dnaX\_2"  
/codon\_start=1  
/transl\_table=11  
/product="DNA polymerase III gamma and tau"  
/protein\_id="BA098482.1"

QY 721 CTTTAACTGCGACTTCGAGATGCTCTTTTAGTAACTCTCTCAACAAGGGA 780  
| | | | |  
Db 12051 CTTTAACTGCGACTTCGAGATGCTCTTTTAGTAACTCTCTCAACAAGGGA 12110  
QY 781 GGAGCAATTCGTACTACAGAGGCGCTCGATACCAATAACAGAGTTATGTTAGATTC 840  
| | | | |  
Db 12111 GGAGCAATTCGTACTACAGAGGCGCTCGATACCAATAACAGAGTTATGTTAGATTC 12170  
QY 841 CTATCTAATCATAGGTCTAGCTCAGAGGCGCTATCGATGATGAAGCAGCTCGATCTA 900  
| | | | |  
Db 12171 CTATCTAATCATAGGTCTAGCTCAGAGGCGCTATCGATGATGAAGCAGCTCGATCTA 12230  
QY 901 TCAGAACAAATTTCTATATTTTGAAGGGAATGAGGAAATCTAGGGGCGGATC 960  
| | | | |  
Db 12231 TCAGAACAAATTTCTATATTTTGAAGGGAATGAGGAAATCTAGGGGCGGATC 12290  
QY 961 TGCAACACCAAGGCGAGTGTATCTCTGAACTGATATCTCTAACAAATAGACTCTGATC 1020  
| | | | |  
Db 12291 TGCAACACCAAGGCGAGTGTATCTCTGAACTGATATCTCTAACAAATAGACTCTGATC 12350  
QY 1021 TTTCTCTCAACGTAGCAGAAACAGCGGTGGCCATCGATGCTTAAAGCTAGCCCTT 1080  
| | | | |  
Db 12351 TTTCTCTCAACGTAGCAGAAACAGCGGTGGCCATCGATGCTTAAAGCTAGCCCTT 12410  
QY 1081 TCCCTGAGGCTTTAGAGAGTTCTACAGAAATATGTCATCAGCAACTCTAAGGGG 1140  
| | | | |  
Db 12411 TCCCTGAGGCTTTAGAGAGTTCTACAGAAATATGTCATCAGCAACTCTAAGGGG 12470  
QY 1141 GGTGCTATCAGCATGATGCTCTCAGAGAGCTCAGTCTTCTGAGAGACAGAAACAT 1200  
| | | | |  
Db 12471 GGTGCTATCAGCATGATGCTCTCAGAGAGCTCAGTCTTCTGAGAGACAGAAACAT 12530  
QY 1201 ACCCTTGTAGAAATACCTTACAAACACCGGAAGTACCGATCTCTTAAACGTATGCG 1260  
| | | | |  
Db 12531 ACCCTTGTAGAAATACCTTACAAACACCGGAAGTACCGATCTCTTAAACGTATGCG 12590  
QY 1261 ATCAACATAGGAATGAAACGGGAAATTCACGAATTAGGGGTGCTTAAAAATCATACAT 1320  
| | | | |  
Db 12591 ATCAACATAGGAATGAAACGGGAAATTCACGAATTAGGGGTGCTTAAAAATCATACAT 12650  
QY 1321 TTTCTTATATGATCCCATCTCTCAGAAAGAACTCATCAGAGATTTAGAAATTAATAC 1380  
| | | | |  
Db 12651 TTTCTTATATGATCCCATCTCTCAGAAAGAACTCATCAGAGATTTAGAAATTAATAC 12710  
QY 1381 GGCCTCTGCGGAGCTCTCAATCCATATCAGAGAACGATTCATATTTCTGGGAAACCTTA 1440  
| | | | |  
Db 12711 GGCCTCTGCGGAGCTCTCAATCCATATCAGAGAACGATTCATATTTCTGGGAAACCTTA 12770  
QY 1441 ACAGCAATGAACTTAAAGTGTGACAAATTTAAATCTTCAATCAGCAGCAGCTCTCC 1500  
| | | | |  
Db 12771 ACAGCAATGAACTTAAAGTGTGACAAATTTAAATCTTCAATCAGCAGCAGCTCTCC 12830  
QY 1501 CTATCCGAGAGAAATGATTCCTACAAAAGGAGTCACTTTAGAGAGCAGAGCTTCTCT 1560  
| | | | |  
Db 12831 CTATCCGAGAGAAATGATTCCTACAAAAGGAGTCACTTTAGAGAGCAGAGCTTCTCT 12890  
QY 1561 CAAAGAGCGGCTTCTCTCTGCGATGATTCAGAGAACGATTCATCACTACAGCTGCG 1620  
| | | | |  
Db 12891 CAAAGAGCGGCTTCTCTCTGCGATGATTCAGAGAACGATTCATCACTACAGCTGCG 12950  
QY 1621 AGTTTTCATCATCAGAACCTAGGAATCAATGTTGACTCCTTAGTCTTAAACAGCCGCTC 1680  
| | | | |  
Db 12951 AGTTTTCATCATCAGAACCTAGGAATCAATGTTGACTCCTTAGTCTTAAACAGCCGCTC 13010  
QY 1681 AGCTTAACAGCAAAAGGTGCTTCAATTAAGTGAATGATGATGGAAGCTCAACCTGAT 1740  
| | | | |  
Db 13011 AGCTTAACAGCAAAAGGTGCTTCAATTAAGTGAATGATGATGGAAGCTCAACCTGAT 13070  
QY 1741 GATATTTGAAGGAACATTTATGAAGTCAATATGTTTCAGCCATGACAGAGCTTCTCTCTA 1800  
| | | | |  
Db 13071 GATATTTGAAGGAACATTTATGAAGTCAATATGTTTCAGCCATGACAGAGCTTCTCTCTA 13130

QY 1801 TTAATAATCAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860  
| | | | |  
Db 13131 TTAATAATCAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13190  
QY 1861 GTTCTGCTGAGAGTCTTAATTCAGAAATAGGATTCAGAGCAATGGAATGTTAATGG 1920  
| | | | |  
Db 13191 GTTCTGCTGAGAGTCTTAATTCAGAAATAGGATTCAGAGCAATGGAATGTTAATGG 13250  
QY 1921 ACTACGATACAGCTACAAATACAAAGAGGCGCAGGCACTTGGAGCCAAACAGGATTT 1980  
| | | | |  
Db 13251 ACTACGATACAGCTACAAATACAAAGAGGCGCAGGCACTTGGAGCCAAACAGGATTT 13310  
QY 1981 GTTCCACAGCCCGGAAAGAAATTCGCTTACTATATGCAATATACCTATGAGAGTCTTACT 2040  
| | | | |  
Db 13311 GTTCCACAGCCCGGAAAGAAATTCGCTTACTATATGCAATATACCTATGAGAGTCTTACT 13370  
QY 2041 GACATTCGCTCTCTGCAACAGCTTGTAGAGATCGGCGCACTGATGGAACACAAACAA 2100  
| | | | |  
Db 13371 GACATTCGCTCTCTGCAACAGCTTGTAGAGATCGGCGCACTGATGGAACACAAACAA 13430  
QY 2101 GGTTCGAGGTTTCTCCATGACGAACCTTCTGCATTAAGTACAGTGAATGGAATCGCAAA 2160  
| | | | |  
Db 13431 GGTTCGAGGTTTCTCCATGACGAACCTTCTGCATTAAGTACAGTGAATGGAATCGCAAA 13490  
QY 2161 GGTTCGAGGTTTCTCCATGACGAACCTTCTGCATTAAGTACAGTGAATGGAATCGCAAA 2220  
| | | | |  
Db 13491 GGTTCGAGGTTTCTCCATGACGAACCTTCTGCATTAAGTACAGTGAATGGAATCGCAAA 13550  
QY 2221 GACCTATTTACCTTTCGCTTCGCATCTCTTTCGTAGAGACAAAGATTTGTTATGCT 2280  
| | | | |  
Db 13551 GACCTATTTACCTTTCGCTTCGCATCTCTTTCGTAGAGACAAAGATTTGTTATGCT 13610  
QY 2281 CACAAACACCTAGAACCTAGGAGGAACTTATTTCTCAAGCATCTCATACCTTACAA 2340  
| | | | |  
Db 13611 CACAAACACCTAGAACCTAGGAGGAACTTATTTCTCAAGCATCTCATACCTTACAA 13670  
QY 2341 CCCCCAAACATTTAGATTTAGAAAGCAAAAGTTTCTGAAATAGATATGAAAAATTC 2400  
| | | | |  
Db 13671 CCCCCAAACATTTAGATTTAGAAAGCAAAAGTTTCTGAAATAGATATGAAAAATTC 13730  
QY 2401 CCTAGGAAATTCCTTACAGCTTGTGATGTCAGATTTTCCTTCAACCCATCTCTTCAAG 2460  
| | | | |  
Db 13731 CCTAGGAAATTCCTTACAGCTTGTGATGTCAGATTTTCCTTCAACCCATCTCTTCAAG 13790  
QY 2461 ATGGAAGGCACTATACCTCATATGCGAATTCGGAAGTTTGGAGCAAGAGTATTA 2520  
| | | | |  
Db 13791 ATGGAAGGCACTATACCTCATATGCGAATTCGGAAGTTTGGAGCAAGAGTATTA 13850  
QY 2521 GCTGCTGATACGCGCTAGACCTTCTTGTCTTCTTCAACCCATCTCTTCAAG 2580  
| | | | |  
Db 13851 GCTGCTGATACGCGCTAGACCTTCTTGTCTTCTTCAACCCATCTCTTCAAG 13910  
QY 2581 ACCCTTCAATTCACAGATGAAAGTGAATGTTATGATACAAATAGCTTCTTGAA 2640  
| | | | |  
Db 13911 ACCCTTCAATTCACAGATGAAAGTGAATGTTATGATACAAATAGCTTCTTGAA 13970  
QY 2641 AGCTCTAGTATGCGCGGTTTATGATTTGAAAGGCGCTTAACTCTGATTCCTG 2700  
| | | | |  
Db 13971 AGCTCTAGTATGCGCGGTTTATGATTTGAAAGGCGCTTAACTCTGATTCCTG 14030  
QY 2701 GGTGCGAAATTCGTGACAGGAGGATTCGGAATTCCTACACTATGATCTCAGATTC 2760  
| | | | |  
Db 14031 GGTGCGAAATTCGTGACAGGAGGATTCGGAATTCCTACACTATGATCTCAGATTC 14090  
QY 2761 TTTGTTTCCATGCTATCTATCAATCCCAATCTACAGGACTCTTGTATGAGCCCA 2820  
| | | | |  
Db 14091 TTTGTTTCCATGCTATCTATCAATCCCAATCTACAGGACTCTTGTATGAGCCCA 14150  
QY 2821 GACTCTTGAAGAAATTCGCGGTCGAATCTTTCAGACAGGATTTTACTAGGGGTAC 2880  
| | | | |  
Db 14151 GACTCTTGAAGAAATTCGCGGTCGAATCTTTCAGACAGGATTTTACTAGGGGTAC 14210  
QY 2881 AACAACTACGCTACACTCCAAATTTGAGCTCTTGGACATTTAGGATGGAACCTCGT 2940

CDS complement(6954..8060)  
/note="ORF1"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAB37070.1"  
/db\_xref="GI:4455885"  
/db\_xref="SPTREMBL:O92897"  
/translation="MSPQOPEOCHFDVGSFLRPESLTRASDFEGRVYEQMRV  
ENAIRNLKOTENGILFTTDGERRRRSMDEMGHGDGRDSDNDPEIGVILKD  
KTSVSKHPETHEFEVKTEKGNNAKOTIPSPSQFHEHMLFAPNLKMTKRYENOE  
LDDIVFYROYVLODLVLAAGCNLQDLCAMCRLLDIRAPSWYGVSDSHRLHLEOF  
LMIHLMVKMREDREDFVSLHVCGRDYOAEFFSRAYDSIEEPLFAKTVDVSHYEWAL  
DKYSGAEPLAYSGEKHVCGLGLISSNCSIEDDVAVSRLEYASTIPERLSTSP  
OCGFAACGSDHMTPEEBOKKIAFYKEIAKEIMWG"  
complement(8586..11372)  
/gene="omp5"  
complement(8586..11372)  
/gene="omp5"  
/codon\_start=1  
/product="outer membrane protein 5"  
/protein\_id="CAB37071.1"  
/db\_xref="GI:4455886"  
/db\_xref="SPTREMBL:O9RB65"  
/translation="MKSQFQMLVLSSTLACFTSCSTVFAATENIGPSDFGNTNG  
TTPKNTTGTIDYTLTGDTTLQNLGDSALATKQCPDSTTESLSPAGKYSLEPLKIS  
SABGAALSVTDDKNLSLTFGSSVTLFLAPSSVITTPSGKAGKCGDILFDPNGLTIF  
KODICEENGAGISTNLSLKNSTGSISEGNKSSATGKGGAICTATGVDITNNAPT  
LEFNNIAAGAGAINSTGNCITTGNTLSIFESNSVATGKGGAICTATGVDITNNAPT  
VTFESQNAVANGAIYAKKILTLASGGGGISFNNIYOGTTANGAALSILAAGCESL  
SAEADITFENGNAIVATTPOPTKRNISIDIGSPAKITNLRATGHSIFPYDPTATTA  
DSTDLTNKADAGNSTDYSISIVFSGKLSDEKAVADNLSTIKOYVTLTAGNLV  
KRGVTLDKGTQTAGTSSVIMAGTTAKASFEETVLTGLSTIPVDSLGSKYVNA  
ASKNALSGPILILDNNGNAIENHDLGTQDSEYQLSALGATTTDVPAYTPVPT  
HIGVGTGEMTWDDVDASTPKTKATATLMTNTGYLPNERQDPLVPSLMSQSDIOA  
IUGVTERALTLCDRFGWAGVANPLDKMTNGEGRKRYRHSGGAIAGAGQCESL  
ISFAEQLEFGSKDLELVAKNHDIYAGAFYIOHITECSEFICGLDKLPGSSHKLPL  
LEGOLAYSHVSNDELTKTATAYEVKSGNNAFNNGLASSHSYEPGLPCTVAPYI  
KLNTYIRDSFSEKTEGSRFSDSNLNLNLSIPGKREKESDCDPSYDLTLISVDP  
LIRNDPKCTTALVIGASAMETRYANLALQALQVRGSHYASPMVEYLGQFVFEVRS  
SRINVDLGKRFQF  
complement(11313..11372)  
/gene="omp5"  
complement(11313..11372)  
/gene="omp5"  
/gene="omp4"  
11535..14321  
11535..11588  
/gene="omp4"  
11535..14321  
/gene="omp4"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein 4"  
/protein\_id="CAB37072.1"  
/db\_xref="GI:4455887"  
/db\_xref="SPTREMBL:O86164"  
/translation="MKTSTPWLIVSSVLAFCSHLOSLANEELSPDSENGINDSGT  
TPKSTATTYSLTGDFEYEPGKGPLNSCRKOTDNLTPFGNGLSPLEPGDTHA  
GAAATYANKNLTSGFSLSDSPSTVTTTGQSTLSAGGVNIENIRKLVAGNES  
TADGAIKIGASFLGTGSDALFNNSSSTGAGLATTAGARLIANTYVFLNLSIAS  
TSGAIDEGESTILSNKFLYFEGNAATGGAICONTASPELIIENNTLIPASN  
VAETSGAIIHAKKILALSSGFTEFLNNVSSATPKGASISIDASELSLAETGNT  
VRNTLTGTSTDPKRNAINIGSNKFTPELRAKWHITFEYDPTISEGSDVLKIN  
GSAGALNYOGTILFSGFTLRADLKVDNLKSSPTQVYSLGGGLLQKQVTLTETS  
FSQEGSLTGMDSGTTLSLTAGSTITTMGLINVDLGLQPSVLRKAGKSNVYEGK  
LNLIDEGNIYESHMSHDLFLSLKITVADVADVNDVLSLIPVAPADPSEKVEGQ  
QMNVMWTDATANTKEATATYTKGTFVSPERKSAVONTLMGVTDRSLQOLVIG  
ATGHEKQGFVWVSMNTFLHKTGDENKRGFRHTSGGYIGSASHTPKDRLTFAPCIG  
FARDQCFIAHNSSTYGTGLFEFKSHLOPONTYLRGAKESBAIEKPPREDPLAL  
DVQVFSHSNDRMETHYSLSPESBSHNECIAGIGLDPVLSNPHDKPTETPDM  
KVENYVYQNSFESSSDGRGFSIGRLNLTSPGAKRYOGDIDGSIYTDLSGFFVSD  
VYRNPOSTATLWSPDSWKIRGNLSHQAFLLKRSNNVYVNSNCELGHVAMELRGS  
SRNVVDVGTGKLRP"

sig\_peptide 14579..14644  
/gene="omp13"  
14579..16123  
/gene="omp13"  
14579..16123  
/gene="omp13"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB37073.1"  
/db\_xref="GI:4455888"  
/db\_xref="SPTREMBL:O923D6"  
/translation="MTILRNFLTCSALFLALPAAQVYVLAHESDQVNGAINKSLEPK  
ITCYPEGTSYIFLDDVLRISNWKHDDQDAGVFINRSGNLFEMGNRNFETHNLTGFG  
AASIRVQDPTLTLSNFSLAFTSAPLIPGOGGALYSIGSVAIENSEEYTFGNSW  
SGAALTYPLIGSKSRSPSVNLSNGRVLFRNVSOGGALISTHNLITLTPGSCPE  
NNHAIHDVNSNGALAIAPGGSISIVSGDILFRGNRASQDGRNTIHSIHQSNOF  
KNLRAVSEGVYFDPISHSESHKITDLVINAPBEKREYEGTISFGCLDHEVCAB  
NLTSYIIDDVTLAGGTLISLSDQVLTLSLSEFKQEAESTLMSPTLLCSGDARVONLH  
ILIEDDNFVPRIRAEKDKALVSLKLVAFEAQVSYDFQFEARTIPLLELWNP  
SPDSLLEGTTLERTQVTTENDAVRGFWSLSMEEYVPSLDKDRITTPRKVYFLWNP  
ETTSRP"

Query Match 100.0%; Score 3200; DB 2; Length 17280;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATCGAAGAGACGACATACAGAAATTCGATTCATTAACCCATTATATATA 60  
|||||  
Db 11331 CAATCGAAGAGACGACATACAGAAATTCGATTCATTAACCCATTATATATA 11390  
|||||  
QY 61 AATCTTACTTCGTCATTTAAATAGAAAACACAGAGTCAGATATAAAATCTTGAC 120  
|||||  
Db 11391 AATCTTACTTCGTCATTTAAATAGAAAACACAGAGTCAGATATAAAATCTTGAC 11450  
|||||  
QY 121 AGCTGTTTGTCACTTTACTTGAATTAATTTGTTCTATATTCATGCGAATAGT 180  
|||||  
Db 11451 AGCTGTTTGTCACTTTACTTGAATTAATTTGTTCTATATTCATGCGAATAGT 11510  
|||||  
QY 181 TCTCTAAAAACAAAAGCATTACATGAAGACTGATTCCTTGAGTTTACTTCCTCC 240  
|||||  
Db 11511 TCTCTAAAAACAAAAGCATTACATGAAGACTGATTCCTTGAGTTTACTTCCTCC 11570  
|||||  
QY 241 GGTGTAAGCTTTCATGTACCTACAGTACATGACATACAGAGAACCTTATACCTGAT 300  
|||||  
Db 11571 GGTGTAAGCTTTCATGTACCTACAGTACATGACATACAGAGAACCTTATACCTGAT 11630  
|||||  
QY 301 GATAGCTTAAATGGAATATCGATTGAGAAAGCTTTATCCAAAACCTTCAGCACAACA 360  
|||||  
Db 11631 GATAGCTTAAATGGAATATCGATTGAGAAAGCTTTATCCAAAACCTTCAGCACAACA 11690  
|||||  
QY 361 TATTCCTAACAGAGATGCTCTTTTACAGAGCCGGAAAAGGCATCCCTTATCTGAC 420  
|||||  
Db 11691 TATTCCTAACAGAGATGCTCTTTTACAGAGCCGGAAAAGGCATCCCTTATCTGAC 11750  
|||||  
QY 421 AGTTGTTTAAACCAACCAAGCAATCTTACCTCTTGGGGAAGCGTATAGCTTAAAG 480  
|||||  
Db 11751 AGTTGTTTAAACCAACCAAGCAATCTTACCTCTTGGGGAAGCGTATAGCTTAAAG 11810  
|||||  
QY 481 TTTGGCTTATATGATGCTGCGACTATGACAGTGTCTGTCATCTAACACAGCAATAG 540  
|||||  
Db 11811 TTTGGCTTATATGATGCTGCGACTATGACAGTGTCTGTCATCTAACACAGCAATAG 11870  
|||||  
QY 541 AATCTTACCTCTCAGGGTTTCTTACGAGATTTTGATTCCTCCTGCAACAGCGTT 600  
|||||  
Db 11871 AATCTTACCTCTCAGGGTTTCTTACGAGATTTTGATTCCTCCTGCAACAGCGTT 11930  
|||||  
QY 601 ACTAAGGTCAGAGGAACGCTTCTCAGCAGAGGCGTAAATTTAGAAATATTCGTAAA 660  
|||||  
Db 11931 ACTAAGGTCAGAGGAACGCTTCTCAGCAGAGGCGTAAATTTAGAAATATTCGTAAA 11990  
|||||  
QY 661 CTGTGATGTTGCTGGGAATTTTCTACTGAGATGTGTGAGACTATCAAGAGAGCGTCTTC 720  
|||||  
Db 11991 CTGTGATGTTGCTGGGAATTTTCTACTGAGATGTGTGAGACTATCAAGAGAGCGTCTTC 12050  
|||||







|||||  
Db 4587 TTGGCTTTATAGATGCTGGACATGACAGGCTGTCATCTTCAACACCAATATAG 4528  
OY AATCTACCTTCAGAGGTTTTCTTACTAGTTTGATTTCCCTCCAGCAACGGTT 600  
Db 4527 AATCTTACCTTCAGAGGTTTTCTTACTAGTTTGATTTCCCTCCAGCAACGGTT 4468  
OY 601 ACTACAGCTCAGGGAAGCTTCTCAGCAGAGAGGCTAAATTTAGAAATTTCTTAA 660  
Db 4467 ACTACAGCTCAGGGAAGCTTCTCAGCAGAGAGGCTAAATTTAGAAATTTCTTAA 4408  
OY 661 CTGTGATGTTGCTGGGAATTTTCTTACTGAGATGTTGAGCTATCAAGAGAGCTTTTC 720  
Db 4407 CTGTGATGTTGCTGGGAATTTTCTTACTGAGATGTTGAGCTATCAAGAGAGCTTTTC 4348  
OY 721 CTTTTACGAGCTTCGAGATGCTCTTTTATGAAACACTCTTCATCAACAAAGGA 780  
Db 4347 CTTTTACGAGCTTCGAGATGCTCTTTTATGAAACACTCTTCATCAACAAAGGA 4288  
OY 781 GGAGCAATGCTACTACAGCAGAGCGCTGCATAGCAATACAGAGTTATGTTAGATTTC 840  
Db 4287 GGAGCAATGCTACTACAGCAGAGCGCTGCATAGCAATACAGAGTTATGTTAGATTTC 4228  
OY 841 CTATCTAACATAGCTCTACGTCAGAGCGCTATCGATGATGAAGGCACTGATACTA 900  
Db 4227 CTATCTAACATAGCTCTACGTCAGAGCGCTATCGATGATGAAGGCACTGATACTA 4168  
OY 901 TCGAACAAATTTCTATATTTTGAAGGGAATGCGCAAAACTACGCGGTGCGATC 960  
Db 4167 TCGAACAAATTTCTATATTTTGAAGGGAATGCGCAAAACTACGCGGTGCGATC 4108  
OY 961 TCGAACAAAGCGAGGAGATCTCTGAATCTGATACAAATTAAGACTGATC 1020  
Db 4107 TCGAACAAAGCGAGGAGATCTCTGAATCTGATACAAATTAAGACTGATC 4048  
OY 1021 TTTCGCTTCAAAGCTAGCAGAAACAGCGGTGGCGCATCTTAAAGCTAGCCCTT 1080  
Db 4047 TTTCGCTTCAAAGCTAGCAGAAACAGCGGTGGCGCATCTTAAAGCTAGCCCTT 3988  
OY 1081 TCCCTGAGAGGCTTACAGAGTTTCTACGAAATTAATGTCATACGCACTCCTAAGGG 1140  
Db 3987 TCCCTGAGAGGCTTACAGAGTTTCTACGAAATTAATGTCATACGCACTCCTAAGGG 3928  
OY 1141 GGTGCTATCAGATCGATCCTCAGAGAGCTCAGTCTTTCAGAGAGCAGAGAAACATT 1200  
Db 3927 GGTGCTATCAGATCGATCCTCAGAGAGCTCAGTCTTTCAGAGAGCAGAGAAACATT 3868  
OY 1201 ACCTTTGAAGAAATACCTTACAAACCGGAAAGTACCGATACCTCTAAACGTAAATGCG 1260  
Db 3867 ACCTTTGAAGAAATACCTTACAAACCGGAAAGTACCGATACCTCTAAACGTAAATGCG 3808  
OY 1261 ATCAACATGGAAGTACGAGGAAATTCAGGAAATTAAGGGGCTTAAATCATCAAT 1320  
Db 3807 ATCAACATGGAAGTACGAGGAAATTCAGGAAATTAAGGGGCTTAAATCATCAAT 3748  
OY 1321 TTCTCTATGATCCATCACTTCAGAGAGACCTCATCAGAGCTATTTGAAGTAAATTAAC 1380  
Db 3747 TTCTCTATGATCCATCACTTCAGAGAGACCTCATCAGAGCTATTTGAAGTAAATTAAC 3688  
OY 1381 GGTCTGCGGGAGCTCTCAATTCATATCAAGAGAGATTTCTATTTCTGAGAGAAACCTTA 1440  
Db 3687 GGTCTGCGGGAGCTCTCAATTCATATCAAGAGAGATTTCTATTTCTGAGAGAAACCTTA 3628  
OY 1441 ACAGCAGATGAACTTAAAGTCTGCAATTTAAATCTTCATTCACGAGCAGCTCTCC 1500  
Db 3627 ACAGCAGATGAACTTAAAGTCTGCAATTTAAATCTTCATTCACGAGCAGCTCTCC 3568  
OY 1501 CTATCCGAGGAAAGTATTTGCTACAAAAGGAGTCACTTTAGAGAGCAGAGCTTCTCT 1560  
Db 3567 CTATCCGAGGAAAGTATTTGCTACAAAAGGAGTCACTTTAGAGAGCAGAGCTTCTCT 3508  
OY 1561 CAAGAGCGCGGTTCTCTCTCGGCAATGATTCAGAGAGCACTTATCAACTACAGCTGGG 1620  
|||||

Db 3507 CAAGAGCGCGGTTCTCTCTCGGCAATGATTCAGAGAGCACTTATCAACTACAGCTGGG 3448  
OY 1621 AGTATTACATCAGCAACCTTAGAATCAATGTTGACTCTTAGTGGTTTAAAGCCCGTC 1680  
Db 3447 AGTATTACATCAGCAACCTTAGAATCAATGTTGACTCTTAGTGGTTTAAAGCCCGTC 3388  
OY 1681 ACCCTAACAGCAAAAGCTCTTCAATTAAGTGATGATCTGGGAAGCTCAACCTGATT 1740  
Db 3387 AGCCTAACAGCAAAAGCTCTTCAATTAAGTGATGATCTGGGAAGCTCAACCTGATT 3328  
OY 1741 GATATTGAAGGAACTTTATGAAAGTCAATATGTTTACGCAATGACCACTCTTCTCTA 1800  
Db 3327 GATATTGAAGGAACTTTATGAAAGTCAATATGTTTACGCAATGACCACTCTTCTCTA 3268  
OY 1801 TTAAAAATCAGGGTGAATGCTGATGTTGATACTAACGTTGACATCGACGCTTATCCCT 1860  
Db 3267 TTAAAAATCAGGGTGAATGCTGATGTTGATACTAACGTTGACATCGACGCTTATCCCT 3208  
OY 1861 GTTCTGCTGAGGATCCTTAATTCAGAAATACGAATTCGAAGCAATGAATGTAAATGG 1920  
Db 3207 GTTCTGCTGAGGATCCTTAATTCAGAAATACGAATTCGAAGCAATGAATGTAAATGG 3148  
OY 1921 ACTAGGATACAGCTTACAAATACAAAAGGCGCAGGCAACTTGGACCAAAACAGATT 1980  
Db 3147 ACTAGGATACAGCTTACAAATACAAAAGGCGCAGGCAACTTGGACCAAAACAGATT 3088  
OY 1981 GTTCCAGCGCCGAAAGAAATCTGCGTGTAGTATGCAATACCTATGAGGAGCTTACT 2040  
Db 3087 GTTCCAGCGCCGAAAGAAATCTGCGTGTAGTATGCAATACCTATGAGGAGCTTACT 3028  
OY 2041 GACATTCGCTCTCTGCAACAGCTTGTAGAGATCGCGCAACTGTATGGAACAAACAA 2100  
Db 3027 GACATTCGCTCTCTGCAACAGCTTGTAGAGATCGCGCAACTGTATGGAACAAACAA 2968  
OY 2101 GGTTCCTGGGTTCTCTCCATGACGAACCTCTGCAATTAACATGGAATTAATGCGCAA 2160  
Db 2967 GGTTCCTGGGTTCTCTCCATGACGAACCTCTGCAATTAACATGGAATTAATGCGCAA 2908  
OY 2161 GGTTCCTGGGTTCTCTCCATGACGAACCTCTGCAATTAACATGGAATTAATGCGCAA 2220  
Db 2907 GGTTCCTGGGTTCTCTCCATGACGAACCTCTGCAATTAACATGGAATTAATGCGCAA 2848  
OY 2221 GACATTTTACCTTTGCGTTCTGCAATCTCTTCTGTAAGACAAAGATTGTTTATCGCT 2280  
Db 2847 GACATTTTACCTTTGCGTTCTGCAATCTCTTCTGTAAGACAAAGATTGTTTATCGCT 2788  
OY 2281 CACAACACTAGAAACCTACGAGTGAATTTATTTTAAAGCAGCTCTATACCTACAA 2340  
Db 2787 CACAACACTAGAAACCTACGAGTGAATTTATTTTAAAGCAGCTCTATACCTACAA 2728  
OY 2341 CCCCAAAATTTTGAATTTAGAAAGCAAGTTTCTGATACAGCTATAGAAAAATTC 2400  
Db 2727 CCCCAAAATTTTGAATTTAGAAAGCAAGTTTCTGATACAGCTATAGAAAAATTC 2668  
OY 2401 CCTAAGGAAATTTCCCTACGCTTGATGTCAGATTTCTGTAAGCCATTCAGACAAACGT 2460  
Db 2667 CCTAAGGAAATTTCCCTACGCTTGATGTCAGATTTCTGTAAGCCATTCAGACAAACGT 2608  
OY 2461 ATGGAAGCACTATACCTCATTTGCCAGAAATCCGAAGTCTTGTGAAGCAAGAGTATA 2520  
Db 2607 ATGGAAGCACTATACCTCATTTGCCAGAAATCCGAAGTCTTGTGAAGCAAGAGTATA 2548  
OY 2521 GGTGATGATGAGGCTTACCTCTTCTTCTTCCAAACCAACATCTCTTTTCAAG 2580  
Db 2547 GGTGATGATGAGGCTTACCTCTTCTTCTTCCAAACCAACATCTCTTTTCAAG 2488  
OY 2581 ACCCTTATTCACAGATGAAGTGAATGTATATGATCACAAAATAGCTTTTTCGAA 2640  
Db 2487 ACCCTTATTCACAGATGAAGTGAATGTATATGATCACAAAATAGCTTTTTCGAA 2428  
OY 2641 AGCTCTAGGATGGCGGTTTATGATTTGAAGAGGCTTAACTCTCGATTCCTGG 2700  
Db 2427 AGCTCTAGGATGGCGGTTTATGATTTGAAGAGGCTTAACTCTCGATTCCTGG 2368  
|||||

```

/protein_id="AAF38159.1"
/db_xref="GI:7189229"
/translation="MIRLCRFYIDANSILNKSTMTKTSIPWLVSVSLAFSCHLST
ANBELSPDSEFNIDSGTFYKTSATITSLGDFVFEYEGKTLPSDSCFOYTDN
LTVLNGHSLTFEFDAGTHAGAASTTANKNLTFSESLSDSSPTSTYTGQGL
SAGGVLENIRKLVAGNFTADGAIKASFLTGSGALSSNNSTKGAITCT
TAGARINANTGYVFLSNASTSGAIDDEGTILSNKFLYEGNNAKATGCAICHT
KASGPILLISNNKTLIFASNVAETSGAIIHAKKILASGGTFEFLNNVSSAPKGC
AISIDAGSELISAEQNTITFVNTLTSTSTPPKRNATINCSNCKTFPRAKNNT
IFPYDPTSEGTSSDVLKINNGSAGALNPYOGTILRSGELTIDELKVALNLSFQ
PVLISGKLLQKGVLESTFSQEGSLGMDSTLSTYTAGSTITTNLGINVDSIG
LKOPVLTAKGASNKYIVSGKLNLDIEGNIYESHMSHDQSLKLTVDADVATV
DISLIJVPADENSEYEGFQGNVNTTDTATNTEKATATWTKGFPVPSKSAIY
CNTLMGFETDIRSLQOLVEIGATGEMKEGFWVSMNTEHKTGDERKFRHTSGG
VIGSAGTAPKPDLLTFAPACFPARDKCFIAHNSNRVYGTGLPKHSHHTLOPNTLRL
GRKFSBSAIEKPRELPLALDVQVSPSHDNRETHYISLPESGSGWECIAGGIG
LDLPVLSNPHLPKTFEIPQMKVEMYVSONSFESSDGRFSIGRLNLSIPVGA
FVQDIDGSTYDLSGFVSDVYRNPNOSTATLVMSPSWKIRGNLIRQAFILRGSN
NYYNSNCELFGHYAMELRGSSRNINVDGTKLRF"
5026..7812
/gene="CP0303"
/size="7812"
/notes="similar to GP:4376729; identified by sequence
similarity: putative"
/codon_start=1
/translation="polymorphic membrane protein G family"
/product="polymorphic membrane protein G family"
/protein_id="AAF38160.1"
/db_xref="GI:7189230"
/translation="MKSQFSLVLSSTLACFTSCSYFAATAENIGPSDFSDSTNG
TYPKMTTIDYTLTGDIRLONLGSAAITLKGFSPTTESLSPAGKYSLSLNKRS
SAGCALSVTTDKNLSLTGSSSLPLAAPSSTYTPGKCAVACGGLTPDNNGTIF
KQDYCBENGAISTKLSLNKSTGSIISFEBKSSATGKKAICATGTDITNNAPT
LFENNIAEAGGAINSTGCTTGNTSLVESVATAGNGALSGDADVTISGWS
VPSGNOAVANGAIIYAKLTLASGGGGISFNNTVOGTTAGNGAIIILAGESL
SAGDITFNGAIIYATTPOTKRNISIDISTAKITNRLISGHSIFPDYPTANRA
DSQDTLNLANKADAGNSDYSIVFSEKSLSEDAVADNLSTLQOPVLTANVTL
KRGVTLDTKGFOTAGSSVMDAGTTLKASTREVTGLGISIPDSIGCKVYIASA
ASKNVALSGPILLDNOGATYENHDLAKTDFESVQSLAGITTDVRAVPTVATP
HGVQGTWCTWVDASTPRTKATLAWNTYGLPRLPERQGLVPSLMSGSSDIOA
IOGVIERSALTLTSDRGFMAAGVAFNLKDKKGEKRRKSGVYALGGAQCSSENL
ISFAPQLFSGSDPLVAKNHTDYAGAFYIOHTTECSGILPELKLPGSMWKPIY
LEQOLAYSHVSNLKTKEYAYPEVKGSGNNAFMMGASGSHSYLPELHCFDYAPYI
KLNLTYYIRDOSESEKGESESPDSNLNFSLIPLGKFEKFSQCNPSYDITLSTYD
LIRNDPKCTAIIYISGASWETTYANNLARAQLQVRAASHYAFSPMEVLQGFVEVAGS
SRITNVDLGGKPOF"
complement(8170..8292)
/gene="CP0304"
/size="8292"
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/translation="hypothetical protein"
/product="hypothetical protein"
/protein_id="AAF38161.1"
/db_xref="GI:7189231"
/translation="MDLSRWLNKLKTRQLDONIKYDKCSYCLRGISQILGNV"
8329..9444
/gene="CP0305"
/size="9444"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/translation="conserved hypothetical protein"
/product="conserved hypothetical protein"
/protein_id="AAF38162.1"
/db_xref="GI:7189232"
/translation="MEMSPFQPEQCHFDVGSFLRPESLTPRARSDFEGRITVBO
RVVEDAIIENLRIKQTEAGLIFETDEGFYSWDPFMNGFHCVDRRRSINDEIYV
LKDKISVSKHPFIEHEFEVKTPEKNAKAKQITIPSPQFHEMIFAPNLKNTKFPY

```

```

NOELIDIVEYRVOYIIDLTYAACRNLQDLCAMCRLLDIRASWYGVSDHRLQEL
EOLFATLHNLKDRPEDLFVSLHVCNRDYOAEFSESRASVSIPELFAKDYVSHY
WALDXYSGGAPPLAYSGEKHYCLGISNHCIDROAVSRITEASVYILERLS
LSPOQGCASCSEGHRTMEORKKIAFVKETAKEING"
gene
complement(9785..12571)
/gene="CP0306"
/size="12571"
/notes="similar to GP:4376731; identified by sequence
similarity: putative"
/codon_start=1
/translation="polymorphic membrane protein G family"
/product="polymorphic membrane protein G family"
/protein_id="AAF38163.1"
/db_xref="GI:7189233"
/translation="MKSLSWFLISSLALPLSLNFSAPVAVINLGPNSGCT
YTPADQTTNDGTITNLTGVSITNMGSPALASCFKETTGLSOGHOYPLQNI
DAGANCFYTPANKLISFSGFYSILQITNATTTGTAIKSGASIOSNYSCEYGO
NFSNDNGALOGSSISLSLNPILFAKNKATOKRGALYSTGKITNNTLNSASFENT
AANNGATYEASSTISSNKATSEFINNSVATSAAGAIYCSSTSAKPKVLTLDNGE
LNFIGNATISGCAITTDNLVLSGGPTLFFKNNSAIDTPARLOGALAIADSGSLSA
LGSDITFEGNTVYKASSQTTNRNSINIGNTAKIVQLASGNTITFDLPITTSIT
LAUSDALNLNMGPDLAGNPAYOGTIVFSGEKLSEAEADNLKST10QPTLLAGOLS
LKSGVTLVAKSFESQSPSTLMDAGTTLTADGITTNNLVANDYSLEKTKAKLAKAO
ASQTVLSSGSLSDPSGANYEDVSNMNPQVFCCLTLPADDPANIHITDLADPLEKN
PIHMGOGNVALSMOEDTATKSKAATLTWTGYNPNPERGTLVANITGSPVDVRS
IOOLVATKROSOETETGICEGINSFPHKDSITKNGFRISAGVYVGAATTLASDNL
ITAFQOLREGKDRDHFITNKRAVAVASLHLOHLALSSLSLRYLPSSSEDPVLPD
AOISLYSKNTMTKTYTTOAPKSGESSWNBDCALSLASLPHLTLSEGHFAFPIK
VEASYIHODSEKERNNTLVAFSDGDLINVSIVGITFEFSESNERASVATYIYAD
VYKRNPDCTTALLINIMTKTGTNLSRQAGIGRAGIFVAFSPNLVETSNLSMEIRGS
SRSYNVDLGGKPOF"
BASE COUNT 3580 a 2439 c 2890 g 3767 t
ORIGIN
Query Match 100.0%; Score 3200; DB 1; Length 12676;
Best local similarity 100.0%; Pred. No. 0;
Matches 3200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATGCGAAGAGAGCACTAACAGAAATTCGGATTTCTATAACCCACTTATATTA 60
Db 5067 CAATGCGAAGAGAGCACTAACAGAAATTCGGATTTCTATAACCCACTTATATTA 5008
QY 61 AATCTTACTTGGCTCATATTAATACAAAACACAGAGAGCAAGATTAATAATCTTGAC 120
Db 5007 AATCTTACTTGGCTCATATTAATACAAAACACAGAGAGCAAGATTAATAATCTTGAC 4948
QY 121 AGCTGTTTTCATCTTAACTTAACTTATTTGTTTCATATTTGATGCGAATAGT 180
Db 4947 AGCTGTTTTCATCTTAACTTAACTTATTTGTTTCATATTTGATGCGAATAGT 4888
QY 181 TCTCTAAAAAACAAGACATTCATAGACATTCGATTCCTGGGTTTATGTTCTCC 240
Db 4887 TCTCTAAAAAACAAGACATTCATAGACATTCGATTCCTGGGTTTATGTTCTCC 4828
QY 241 GTGTACTTCTTCATGTCACATTCACATTCACATTCACAGAGAACTTTATACCGTAT 300
Db 4827 GTGTACTTCTTCATGTCACATTCACATTCACATTCACAGAGAACTTTATACCGTAT 4768
QY 301 GATAGCTTTTAATGAATATCATTCATGAGAACGTTTACTCCAAAACCTTGAGCCACACA 360
Db 4767 GATAGCTTTTAATGAATATCATTCATGAGAACGTTTACTCCAAAACCTTGAGCCACACA 4708
QY 361 TATTCCTTAACAGAGATGCTTCTTTTACAGAGCTTGAAAAGGCACTCCCTATCTGAC 420
Db 4707 TATTCCTTAACAGAGATGCTTCTTTTACAGAGCTTGAAAAGGCACTCCCTATCTGAC 4648
QY 421 AGTTGTTTAAAGCAACACAGACATTCATCTTCTTGGGAAAGGCTCATAGCTTAACG 480
Db 4647 AGTTGTTTAAAGCAACACAGACATTCATCTTCTTGGGAAAGGCTCATAGCTTAACG 4588
QY 481 TTTCGCTTTATAGATGCTGACACTCATGAGGTCTGCTGCATCTTACACAGCAATAAG 540

```

```

OY 2281 CACAACTAGAACCTACGCGGAACTTTATTTCTCAGCACTCTCATACCTACAA 2340
    |||||||
Db 4981 CACAACTAGAACCTACGCGGAACTTTATTTCTCAGCACTCTCATACCTACAA 5040
OY 2341 CCCCCAACTATTTGAGATTAGAGAGCAAGTTTCTGAAATCAGCTATAGAAAAATTC 2400
    |||||||
Db 5041 CCCCCAACTATTTGAGATTAGAGAGCAAGTTTCTGAAATCAGCTATAGAAAAATTC 5100
OY 2401 CCTAGGGAATTTCCCTACCTTGATGTCAGAGTTTCTTACGACATTCAGACAACTC 2460
    |||||||
Db 5101 CCTAGGGAATTTCCCTACCTTGATGTCAGAGTTTCTTACGACATTCAGACAACTC 5160
OY 2461 ATGGAACGACATATACCTATTCAGAGATTCAGAGATTCGAGACAGAGTATATA 2520
    |||||||
Db 5161 ATGGAACGACATATACCTATTCAGAGATTCGAGAGATTCGAGACAGAGTATATA 5220
OY 2521 GCTGATGATGCGGCTAGACCTTCCTTTGTTCTTTCACACCACTCTCTTTCAAG 2580
    |||||||
Db 5221 GCTGATGATGCGGCTAGACCTTCCTTTGTTCTTTCACACCACTCTCTTTCAAG 5280
OY 2581 ACCTTATTCACAGATGAAAGTCGAAATGCTTTATGATCAGAAATAGCTTTTCGA 2640
    |||||||
Db 5281 ACCTTATTCACAGATGAAAGTCGAAATGCTTTATGATCAGAAATAGCTTTTCGA 5340
OY 2641 AGCTAGTATGAGCGCGGCTTTAGTATTTGAAAGGCTCTTAACCTTCGATTCCTG 2700
    |||||||
Db 5341 AGCTAGTATGAGCGCGGCTTTAGTATTTGAAAGGCTCTTAACCTTCGATTCCTG 5400
OY 2701 GGTGCGAAATTCGTGAGGAGGATATCGAGATTCCTACACCTATGATCTCAGAGATTC 2760
    |||||||
Db 5401 GGTGCGAAATTCGTGAGGAGGATATCGAGATTCCTACACCTATGATCTCAGAGATTC 5460
OY 2761 TTTGTTTCGATGCTATATGTAACAATCCCCAATCTACAGGAGCTTTGATGAGCCCA 2820
    |||||||
Db 5461 TTTGTTTCGATGCTATATGTAACAATCCCCAATCTACAGGAGCTTTGATGAGCCCA 5520
OY 2821 GACTCTTGAAAAATTCGCGGTGGCAATCTTTCAGAGAGGAGGATTC 2880
    |||||||
Db 5521 GACTCTTGAAAAATTCGCGGTGGCAATCTTTCAGAGAGGAGGATTC 5580
OY 2881 AACAACTAGCTACAACTCCAAATGTTGAGCTCTTTCGACATTAGCGTATGGAATCCGT 2940
    |||||||
Db 5581 AACAACTAGCTACAACTCCAAATGTTGAGCTCTTTCGACATTAGCGTATGGAATCCGT 5640
OY 2941 GGAATCTCAAGAACTACGATGATGATGTTGATACCAACTCCGATTCAGATTCGTAA 3000
    |||||||
Db 5641 GGAATCTCAAGAACTACGATGATGATGTTGATACCAACTCCGATTCAGATTCGTAA 5700
OY 3001 ACTCCCTAGCTCTTCTAGGAGATTTTCTCATCTTTTACGAGAAATTTGCTATAGGAA 3060
    |||||||
Db 5701 ACTCCCTAGCTCTTCTAGGAGATTTTCTCATCTTTTACGAGAAATTTGCTATAGGAA 5760
OY 3061 TCGTTTCTCTGCAAACTGTAAAAAATACATTTGTCCTCTTCAAAAAAGATTTCTTTA 3120
    |||||||
Db 5761 TCGTTTCTCTGCAAACTGTAAAAAATACATTTGTCCTCTTCAAAAAAGATTTCTTTA 5820
OY 3121 ATTAATTTCTAGTTATTAATTTTAAAAACAGTTAAATTAATTAAGACAATATCTTA 3180
    |||||||
Db 5821 ATTAATTTCTAGTTATTAATTTTAAAAACAGTTAAATTAATTAAGACAATATCTTA 5880
OY 3181 TTCTATATGACTCTTTT 3200
    |||||||
Db 5881 TTCTATATGACTCTTTT 5900

```

```

RESULT 3
AE002192/c 12676 bp DNA BCT 30-MAY-2000
LOCUS Chlamydia pneumoniae AR39, section 25 of 94 of the complete
DEFINITION genome.
ACCESSION AE002192 AE002161
VERSION AE002192.2 GI:163405

```

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
gene
CDS
gene
CDS

```

Chlamydia pneumoniae AR39.  
Chlamydia pneumoniae AR39.  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

1 (bases 1 to 12676)  
Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Ulteback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouli, H., Craven, B., Bowman, C., Dodson, R., Gwin, M., Nelson, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.  
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
Nucleic Acids Res. 28 (6), 1397-1406 (2000)

2 (bases 1 to 12676)  
Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Ulteback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouli, H., Craven, B., Bowman, C., Dodson, R., Gwin, M., Nelson, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.  
Direct Submission  
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
On Jun 1, 2000 this sequence version replaced gi:7189226.  
Location/Qualifiers  
1. 12676  
/organism="Chlamydia pneumoniae AR39"  
/strain="AR39"  
/db\_xref="taxon:115711"  
/note="synonym: Chlamydia pneumoniae AR39"  
complement(90. 185)  
/gene="CP0300"  
complement(90. 185)  
/gene="CP0300"  
/note="hypothetical protein; identified by glimmer2; putative"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AF38157.1"  
/db\_xref="gi:7189227"  
/translation="MTSLGRSAIFFTGSLKHLRGLFELFOK"  
complement(275. 1819)  
/gene="CP0301"  
complement(275. 1819)  
/gene="CP0301"  
/note="similar to GP:4376735; identified by sequence similarity; putative"  
/codon\_start=1  
/trans\_table=11  
/product="polymorphic membrane protein G family"  
/protein\_id="AF38158.1"  
/db\_xref="gi:7189228"  
/translation="MTILRNLTCGALFLAPAAQVYLHESDYGNAINKSLEPKITCYEGTSYFLFDVRISNVKHDGAVINRGNLFFMGNRGNPFPHNLMGFGAAISNRVGDFTLTLSNFSYLAFTSAPLIPGCGAITSIGSVIERSEFTGNSVSGAALYRPTLYLIGSKASRPSVNLSGNRYLFPDONSOGGAISTHNLTLTTRGSCFENHAYHDVNSGALAIAPGSGISISYVSGDLIRGNASODGNTIHSIHQSAOPNLRVAVSGYVFDPISSHESHKTTDLVINAPGKEYEGTISFGCLDDHEVCALNLTSTLIDVTLAGGTLSDGVTQLHSPQEAASSTLMSPGTLLSGARVONLILIEDTDFEVRIRAEKDALVSLERKLVAFEAWSYVDFPQFEATPIPLELIGSFSDSLILGETLERTQVTTENDAVRGFWSLSMEETPSLDORDRITPRKVFYFLWNPRTSTP"  
complement(2077. 4926)  
/gene="CP0302"  
complement(2077. 4926)  
/note="similar to GP:4376733; identified by sequence similarity; putative"  
/codon\_start=1  
/trans\_table=11  
/product="polymorphic membrane protein G family"

|||||  
Db 2761 AATCTTACTTGGCTCATATTAATAATAGAAAACCTAGAGAGCAAGATATAAATCTTGAC 2820  
OY 121 AGCTGTTTTGTCATCTTTAACTGATTTACTTATTTTGTTCATATTTGATGCAATAGT 180  
Db 2821 AGCTGTTTTGTCATCTTTAACTGATTTACTTATTTTGTTCATATTTGATGCAATAGT 2880  
OY 181 TCTCTAAAAAACAAGCATTTACATGAAGACCTGCAATCTCTGGGTTTGTGTTTCTCC 240  
Db 2881 TCTCTAAAAAACAAGCATTTACATGAAGACCTGCAATCTCTGGGTTTGTGTTTCTCC 2940  
OY 241 GTGTTTACCTTTCTCATGTCACCTACAGTCACTACATCAAGCAAGAACTTTTATCACTGAT 300  
Db 2941 GTGTTTACCTTTCTCATGTCACCTACAGTCACTACATCAAGCAAGAACTTTTATCACTGAT 3000  
OY 301 GATAGCTTTAATGGAATATGCAATTCAGGAACGTTTACTCCAAAACCTTCAGCCACAACA 360  
Db 3001 GATAGCTTTAATGGAATATGCAATTCAGGAACGTTTACTCCAAAACCTTCAGCCACAACA 3060  
OY 361 TATTCCTTAACAGAGATGCTCTTTTACAGGCTGGAAAAGGCACTCCCTTATCTGAC 420  
Db 3061 TATTCCTTAACAGAGATGCTCTTTTACAGGCTGGAAAAGGCACTCCCTTATCTGAC 3120  
OY 421 AGTTGTTTAAAGCAAAACACGAGCAATCTTACCTTGGGGAACGGTCATAGCTTAACG 480  
Db 3121 AGTTGTTTAAAGCAAAACACGAGCAATCTTACCTTGGGGAACGGTCATAGCTTAACG 3180  
OY 481 TTTGGCTTTATAGATGCTGGACATCATGACAGTGCTGCTGATCTACACACAAATAG 540  
Db 3181 TTTGGCTTTATAGATGCTGGACATCATGACAGTGCTGCTGATCTACACACAAATAG 3240  
OY 541 AATCTTACCTTCTAGAGGTTTCTCTTACTAGTGTGATTCCTCTGACACAAAGGTT 600  
Db 3241 AATCTTACCTTCTAGAGGTTTCTCTTACTAGTGTGATTCCTCTGACACAAAGGTT 3300  
OY 601 ACTACAGGTACAGGAAACGCTTCTCTCAGCAAGAGGCGTAAATTTAGAAATATTCGTAA 660  
Db 3301 ACTACAGGTACAGGAAACGCTTCTCTCAGCAAGAGGCGTAAATTTAGAAATATTCGTAA 3360  
OY 661 CTGTGATGCTGGGAATTTTCTACTGACAGATGGTGAGCTATCAAGAGAGCTCTTC 720  
Db 3361 CTGTGATGCTGGGAATTTTCTACTGACAGATGGTGAGCTATCAAGAGAGAGCTCTTC 3420  
OY 721 CTTTTAACTGACACTTCTGAGATGCTCTTTTGTAAACAACCTTGCATCAACAAAGGA 780  
Db 3421 CTTTTAACTGACACTTCTGAGATGCTCTTTTGTAAACAACCTTGCATCAACAAAGGA 3480  
OY 781 GGAGCAATGCTACTACAGCAAGGCGCTCGCATAGCAATAACACAGGTTATGTAGATTC 840  
Db 3481 GGAGCAATGCTACTACAGCAAGGCGCTCGCATAGCAATAACACAGGTTATGTAGATTC 3540  
OY 841 CTATCTAACAATAGGCTGCTAGCTCAGAGGCGCTCTGATGATGAAGCAGCTGATCTA 900  
Db 3541 CTATCTAACAATAGGCTGCTAGCTCAGAGGCGCTCTGATGATGAAGCAGCTGATCTA 3600  
OY 901 TCGAACAACAATTTCTATATTTTGAAGGGAATGACGAAATCTAGTGGCGTGCAATC 960  
Db 3601 TCGAACAACAATTTCTATATTTTGAAGGGAATGACGAAATCTAGTGGCGTGCAATC 3660  
OY 961 TCGAACAACAAGGCGATGATCTCTGAACGTGAATCTCTTAACAATTAAGACTGTATC 1020  
Db 3661 TCGAACAACAAGGCGATGATCTCTGAACGTGAATCTCTTAACAATTAAGACTGTATC 3720  
OY 1021 TTTGCTTCAAAACGTAAGCAAAACAGGCGTGCGCCATCCATGCTAAAAAGCTTAGCCCTT 1080  
Db 3721 TTTGCTTCAAAACGTAAGCAAAACAGGCGTGCGCCATCCATGCTAAAAAGCTTAGCCCTT 3780  
OY 1081 TCTCTGAGAGGCTTACAGAGTTTCTAGGAATTAATGCTCATAGCAACCTCCAAAGGG 1140  
Db 3781 TCTCTGAGAGGCTTACAGAGTTTCTAGGAATTAATGCTCATAGCAACCTCCAAAGGG 3840  
OY 1141 GGTGCTATACGATCGATGCTCAGAGAGCTAGTCTTTCTGAGAGACAGGAACATTT 1200  
|||||

Db 3841 GGTGCTATACGATCGATGCTCAGAGAGCTCAGTCTTTCTGAGAGACAGAAACATTT 3900  
OY 1201 ACCTTTGTAGAAATACCCTTAGACAACACGGGAAGTACCGATACTCTTAAACGTAATGCG 1260  
Db 3901 ACCTTTGTAGAAATACCCTTAGACAACACGGGAAGTACCGATACTCTTAAACGTAATGCG 3960  
OY 1261 ATCAACATAGGAATTAACGGGAATTCACGGAATTTAGGGGCTGCTAAAAATCATACATTT 1320  
Db 3961 ATCAACATAGGAATTAACGGGAATTCACGGAATTTAGGGGCTGCTAAAAATCATACATTT 4020  
OY 1321 TTTCTTATGATCCCATCACTTCTGAGAAGAACCTCATACAGATTTGAAGATTAATAC 1380  
Db 4021 TTTCTTATGATCCCATCACTTCTGAGAAGAACCTCATACAGATTTGAAGATTAATAC 4080  
OY 1381 GGCCTGCGGGAGCTCTCAATCCATATCAAGGAACGATTCATTTTGTGGGAACCCCTA 1440  
Db 4081 GGCCTGCGGGAGCTCTCAATCCATATCAAGGAACGATTCATTTTGTGGGAACCCCTA 4140  
OY 1441 ACAGCAGATGAATTAAGTGTGACAAATTTAAATCTTCATTCACGACGCCAGCTCTCC 1500  
Db 4141 ACAGCAGATGAATTAAGTGTGACAAATTTAAATCTTCATTCACGACGCCAGCTCTCC 4200  
OY 1501 CTATCCGAGGAAGTTATTTGCTACAAAAGGAGTCACTTTAGAGACACGAGCTTCTCT 1560  
Db 4201 CTATCCGAGGAAGTTATTTGCTACAAAAGGAGTCACTTTAGAGACACGAGCTTCTCT 4260  
OY 1561 CAAAGGCGGGTCTCTCTCCGCGATGATTCAGGAACGACATTTATCACTACAGCTGGG 1620  
Db 4261 CAAAGGCGGGTCTCTCTCCGCGATGATTCAGGAACGACATTTATCACTACAGCTGGG 4320  
OY 1621 AGTATTAACAATCAGAACCTAGGAATCAATGTTGACTCTTAAAGCTCTTAAAGCAGCCGCTC 1680  
Db 4321 AGTATTAACAATCAGAACCTAGGAATCAATGTTGACTCTTAAAGCTCTTAAAGCAGCCGCTC 4380  
OY 1681 AGCCTTAACAGCAAAAGTGTCTTCAATTAAGTGTATCTGGAAGCTCAACCTGATTT 1740  
Db 4381 AGCCTTAACAGCAAAAGTGTCTTCAATTAAGTGTATCTGGAAGCTCAACCTGATTT 4440  
OY 1741 GATTTTGAAGGAACATTTATGAAGTCATATGTTACGCCATGACCCAGCTTCTCTCTA 1800  
Db 4441 GATTTTGAAGGAACATTTATGAAGTCATATGTTACGCCATGACCCAGCTTCTCTCTA 4500  
OY 1801 TTTAAATTAACAGGTTGATGCTGATGATTAATCAACTGATCAACACAGCTTATCCCT 1860  
Db 4501 TTTAAATTAACAGGTTGATGCTGATGATTAATCAACTGATCAACACAGCTTATCCCT 4560  
OY 1861 GTTCTGCTGAGATCTTAATTCAGAAATACGGAATTCAGAGCAATGGAATGTTAATGCG 1920  
Db 4561 GTTCTGCTGAGATCTTAATTCAGAAATACGGAATTCAGAGCAATGGAATGTTAATGCG 4620  
OY 1921 ACTACGATACAGCTACAAATTAACAAAAGGCGCACGCACTTGAGCAAAACGAGATTT 1980  
Db 4621 ACTACGATACAGCTACAAATTAACAAAAGGCGCACGCACTTGAGCAAAACGAGATTT 4680  
OY 1981 GTTCCCAAGCCCGCAAAAATATGCTGTTAGTATGCAATACCTTATGGGAGTCTTACT 2040  
Db 4681 GTTCCCAAGCCCGCAAAAATATGCTGTTAGTATGCAATACCTTATGGGAGTCTTACT 4740  
OY 2041 GACATTCGCTCTGCAACAGCTGTTAGAGATGCGCCAACTGCTATGGAACACAAACAA 2100  
Db 4741 GACATTCGCTCTGCAACAGCTGTTAGAGATGCGCCAACTGCTATGGAACACAAACAA 4800  
OY 2101 GGTTCGAGGTTTCTCATAGCAAGTCTCTGATTAAGCTGGAGATGGAATAATCGCAA 2160  
Db 4801 GGTTCGAGGTTTCTCATAGCAAGTCTCTGATTAAGCTGGAGATGGAATAATCGCAA 4860  
OY 2161 GGCCTCCGATACCTTGAAGGCTAGCTATCGTGGGAAGTGTCTACACTCTCAAAAGAC 2220  
Db 4861 GGCCTCCGATACCTTGAAGGCTAGCTATCGTGGGAAGTGTCTACACTCTCAAAAGAC 4920  
OY 2221 GACCTATTTAACCCTTGGGTTCTGCAATCTCTTGTGTAAGACAAAGATTTTATTCGCT 2280  
Db 4921 GACCTATTTAACCCTTGGGTTCTGCAATCTCTTGTGTAAGACAAAGATTTTATTCGCT 4980  
|||||

Db	3001	ACTGCCAGTCTCTCTAGGAGCTTTTCACACACTTTTAGGAAATATTTTGGTATRAGGAA	3060
QY	3061	TGCTTTCTTCGCAACATGTAATAAATAACATTTGTCCCTCTTCAGAAAAGATTTCTTTTA	3120
Db	3061	TGCTTTCTTCGCAACATGTAATAAATAACATTTGTCCCTCTTCAGAAAAGATTTCTTTTA	3120
QY	3121	ATAATTTCTACTTAAATTTTAAATTTTAAACAGTAAATAATTAATGACAATATCTTA	3180
Db	3121	ATAATTTCTACTTAAATTTTAAATTTTAAACAGTAAATAATTAATGACAATATCTTA	3180
QY	3181	TTCTTAATGACTTCTTTTTT 3200	
Db	3181	TTCTTAATGACTTCTTTTTT 3200	
RESULT	2		
LOCUS	CPOMP54	6030 bp DNA BCT 22-JAN-1999	
DEFINITION	Chlamydia pneumoniae omp5 and omp4 genes.		
ACCESSION	AJ001311		
VERSION	AJ001311.1 GI:3255934		
KEYWORDS	omp4 gene; omp5 gene; outer membrane protein 4; outer membrane protein 5.		
SOURCE	Chlamydia pneumoniae.		
ORGANISM	Chlamydia pneumoniae		
REFERENCE	1 (bases 1 to 6030)		
AUTHORS	Kudsen, K.		
JOURNAL	Direct Submission		
TITLE	Submitted (29-AUG-1997) Kudsen K., Department of Medical Microbiology and Immunology, University of Aarhus, The Bartholin building, University of Aarhus, DK-8000 Aarhus C, DENMARK		
REFERENCE	2 (bases 1 to 6030)		
AUTHORS	Kudsen, K., Madsen, A.S., Mygind, P., Christiansen, G. and Birkeland, S.		
TITLE	Identification of two novel genes encoding 97- to 99-kilodalton outer membrane proteins of Chlamydia pneumoniae		
JOURNAL	Infect. Immun. 67 (1), 375-383 (1999)		
FEATURES	99081766		
source	Location/Qualifiers		
gene	1..6030		
CDS	/organism="Chlamydia pneumoniae"		
	/isolate="CDC/CWL-029/YR-1310"		
	/db_xref="taxon:83358"		
	/db_xref="pdx2-18"		
	/clone="B8F3"		
	/clone_1lb="pbluescript"		
	/dev_stage="elementary body"		
	/complement(1..2804)		
	/gene="omp5"		
	/complement(<1..2742)		
	/gene="omp5"		
	/note="putative lipoprotein"		
	/codon_start=1		
	/transl_table=1		
	/product="Outer membrane protein 5"		
	/protein_id="CA04671.1"		
	/db_xref="GI:3255935"		
	/db_xref="SPRMBL:086163"		
	/translation="MKSQFSLVLSLACFTSCSYFAATNENIGSDSPDSTUNG TYPKATTCIDYICGIDITDILQDLSAATATKCSQSPMTESIPACKGYLSPLNIK SAGGALSYTTDKNLSTLRGSSLTFLEAPESVITTBGKCAVKGGLTIDNKTILF KQYCEENGASISTKNLSLNTSGTISFEFGKSSATGKKGCAICATVTVITNNAPT LFSNNIAEAAGAIINSTGCTIGNTSIVSENSVYATAGNGALSGADAVTISGNSQ VFSGQAVANGGAIYAAKLTJLASGGGGISFSNNIVQGTAGNGAISTIAAGECSL SAPAGCITFENGNAIVATTPTTKRNSIDISGTAKITLRLAISGISIFPDITANPA KREVTDLINKADAGNSIDYSGSIVFSEKSLSDCAKADNLSTIKQPVITLGNIVL KSDVTLDFKGTGFTGFASSSVTMDAGTTLKATSEVYTLGLISTIPDLSCEKKVYIASA ASNNALSGPILLDNNGNAIENHDDJAKTDFSVQLSALGATTTDVPVPAVATP HGVIGQSWGTWTDVASTPRTYATLAWNTGYLNPBEROGLVPSNLGSPSDEAL IIOYIGSTWATLTCSDRFGMAAGVANELDKDKGKEKRYRKRSGGYAIGGAAGTCSENL		

	mat_peptide	complement(1. .2697)	/gene="omp5"	/product="outer membrane protein 5"		
	sig_peptide	complement(2698. .2742)	/gene="omp5"			
	RBS	complement(2747. .2752)	/gene="omp5"			
	-10_signal	complement(2777. .2782)	/note="putative"			
	-35_signal	/gene="omp5"	/note="putative"			
	gene	complement(2799. .2804)	/gene="omp5"			
	-10_signal	2863. .5726	/gene="omp4"			
	RBS	2863. .2868	/gene="omp4"			
	sig_peptide	/note="putative"	/note="putative"			
	CDS	2896. .2902	/gene="omp4"			
		/note="putative"	/note="putative"			
		2905. .5691	/gene="omp4"			
		/note="putative lipoprotein"	/note="putative lipoprotein"			
		/codon_start=1	/codon_start=1			
		/transl_table=11	/transl_table=11			
		/product="outer membrane protein 4"	/product="outer membrane protein 4"			
		/protein_id="CAA04672.1"	/protein_id="CAA04672.1"			
		/db_xref="GI:3255936"	/db_xref="GI:3255936"			
		/db_xref="SPTREMBL:O86164"	/db_xref="SPTREMBL:O86164"			
		/translation="MKTSLIPWLVSSVLAIFSCHLQSLANELLSPPDSENGNIDSGTF TPKTSATTVYSILGDFVFEPEPKGTPILSDSCFKQTTDMLEPLMGHSLTFEPDIAGTHA GAASTYANKNLITGDFVFEPEPKGTPILSDSCFKQTTDMLEPLMGHSLTFEPDIAGTHA TADGAILGASFLITGTSGLDLSFNSNSKGGAILATPAGARIANNTGYRFLSNLS TSGAIDDEGSLISNNKFLYFEGNMAKTTGGAILKONTGASPELLISNNKTLIFASN VAEISGAILHAKKILALSSGGFTETLRNNVASLPCMGKATSIDASGELISAEENQITF VRNLTITGSGTDTPKRNAINIGSKRTTELRANKNHTFFYDPTTSEGTSSDYKINN GSAALNPYOGTILFSEGETLPADELKADNLIKSSFTQPSLSGGKLLQGVLEST FSEGASILGMDSGTILSTAGSITITNLGINVDSLQKQPSLTLAAGASANKYIVSGK LNLDELGINIYESHMFSDQLFSLKLTIVADYDVTNDISLLPVPADENSGEFGG OATNNTTDTATNTKREATITKTKGFPVSPERSALVCNTLMGVFTDRIISLOOLEVG ATGNEHKGQFVSMTEPTEFLAKTGDENKGRPHNSGCVYIGGSAHTPDDILPTAFACHL FARKDCFLIAHNSRITGGTLFERKHSITLDPONTLRGRKPFESALIEKPPRELPAL DVOVSFSDNRMETHTYSLPESGSSNSNCLIAGGICGLDLPVLSNHPLEKTFIPDM KVEVAVYSONSFESSSDRGFPISGRLLNISIPVGAFAFVGGIDIGDSTYTLISEGPSVD VYRNPSTATILVWSPDSMKIRGGLNSROAFLLRGSNNYVNSNCELFGHAMELRNS SRNPNVGTGKLR"				
	mat_peptide	2956. .5688	/gene="omp4"	/product="outer membrane protein 4"		
	stem_loop	5698. .5726	/gene="omp4"			
		/note="putative"	/note="putative"			
	BASE COUNT	1760 a 1341 c 1195 g 1734 t				
	ORIGIN					
	Query Match	100.0%; Score 3200; DB 2; Length 6030;				
	Best Local Similarity	100.0%; Pzed. No. 0;				
	Matches 3200; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
1	CAATGTCGAAGAGGCACTAACCGAGAAATGGATTCCTAAACCACTTATTTA 60					
2701	CAATGTCGAAGAGGCACTAACCGAGAAATGGATTCCTAAACCACTTATTTA 2760					
61	AATTCCTACTTGCGCTCAATATAAGAAAACCTCAGAGAGTCAAGATATAATTCCTTGAC 120					

QY 841 CTATCTACATAGCGTCTACGTCAGAGGCGCTATCGATGATGAGACCGTCATCTACTA 900  
Db 841 CTATCTACATAGCGTCTACGTCAGAGGCGCTATCGATGATGAGACCGTCATCTACTA 900  
QY 901 TCGAACACCAATTTCTATTTTGAAGGAGATGAGCGAAACACTGCGGTCGATC 960  
Db 901 TCGAACACCAATTTCTATTTTGAAGGAGATGAGCGAAACACTGCGGTCGATC 960  
QY 961 TCGAACACCAAGGCGAGTGTCTCTGACGATGATGATCTGACATGATGATGATC 1020  
Db 961 TCGAACACCAAGGCGAGTGTCTCTGACGATGATGATGATGATGATGATGATC 1020  
QY 1021 TTTGGCTTCAACGTCAGGAAACAAAGGCGGCGCATGCTCAATGCTAAAGCTGACCTT 1080  
Db 1021 TTTGGCTTCAACGTCAGGAAACAAAGGCGGCGCATGCTCAATGCTAAAGCTGACCTT 1080  
QY 1081 TCTCTGAGGCGTTCACGAGTCTTCTAGCAATATGTCTCATCGCAACTCTTAAGGG 1140  
Db 1081 TCTCTGAGGCGTTCACGAGTCTTCTAGCAATATGTCTCATCGCAACTCTTAAGGG 1140  
QY 1141 GGTGCTATCAGATCAGTGTCTCTGAGAGGCTCACTCTTCTGACAGACAGAAACATT 1200  
Db 1141 GGTGCTATCAGATCAGTGTCTCTGAGAGGCTCACTCTTCTGACAGACAGAAACATT 1200  
QY 1201 ACCCTTGTAAAGAAATACCTTCAACAACACGGAAGTACGGAATCTCAACGTAATGCG 1260  
Db 1201 ACCCTTGTAAAGAAATACCTTCAACAACACGGAAGTACGGAATCTCAACGTAATGCG 1260  
QY 1261 ATCAACATAGAGAGTAAACGGGAAATTCACGGAATTCACGGGCTGTAAATATCATCAATT 1320  
Db 1261 ATCAACATAGAGAGTAAACGGGAAATTCACGGAATTCACGGGCTGTAAATATCATCAATT 1320  
QY 1321 TTTCTCTATGATCCCATCTCAGAGAGAACCTTCATCAGACGATTTTAAAGTAAATAC 1380  
Db 1321 TTTCTCTATGATCCCATCTCAGAGAGAACCTTCATCAGACGATTTTAAAGTAAATAC 1380  
QY 1381 GGCCTGCGGAGGCTCTCAATCCATCAAGAACGATTTCTATTCTGAGAAACCTTA 1440  
Db 1381 GGCCTGCGGAGGCTCTCAATCCATCAAGAACGATTTCTATTCTGAGAAACCTTA 1440  
QY 1441 ACACGAGATGAATTAAGTGTGACAAATTTAAATCTTCATTCACGACGACGCTCTCC 1500  
Db 1441 ACACGAGATGAATTAAGTGTGACAAATTTAAATCTTCATTCACGACGACGCTCTCC 1500  
QY 1501 CTATCCGAGAGAAATTTCTGCTACAAAAGGAGTCACTTTAGAGAGCAGAGCTTCTCT 1560  
Db 1501 CTATCCGAGAGAAATTTCTGCTACAAAAGGAGTCACTTTAGAGAGCAGAGCTTCTCT 1560  
QY 1561 CAAGGCGCGGCTCTCTCTGCGCATGATTCAGGAAGGACATTTCAACTACAGCTGGG 1620  
Db 1561 CAAGGCGCGGCTCTCTCTGCGCATGATTCAGGAAGGACATTTCAACTACAGCTGGG 1620  
QY 1621 AGATTTCAATCAGAACTTGAATGATGATGATGATGATGATGATGATGATGATGATG 1680  
Db 1621 AGATTTCAATCAGAACTTGAATGATGATGATGATGATGATGATGATGATGATGATG 1680  
QY 1681 AGCCTAACAGAAAGGTGCTTCAATTAAGTATGATGATGATGATGATGATGATGATG 1740  
Db 1681 AGCCTAACAGAAAGGTGCTTCAATTAAGTATGATGATGATGATGATGATGATGATG 1740  
QY 1741 GATTTTGAAGGAACTTTATGAAGATCATGTTTACGCGATGACCGACGCTTCTCTCTA 1800  
Db 1741 GATTTTGAAGGAACTTTATGAAGATCATGTTTACGCGATGACCGACGCTTCTCTCTA 1800  
QY 1801 TTTAAATTCAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1860  
Db 1801 TTTAAATTCAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1860  
QY 1861 GTTCTGCTGAGATCTTAATTCAGAAATGAGATTCAGAGACAAATGGAATGTTAATGG 1920  
Db 1861 GTTCTGCTGAGATCTTAATTCAGAAATGAGATTCAGAGACAAATGGAATGTTAATGG 1920

QY 1921 ACTAGGATACAGCTACAAATACAAAAGAGGCGACGGCACTTTGGACCAACAGATTT 1980  
Db 1921 ACTAGGATACAGCTACAAATACAAAAGAGGCGACGGCACTTTGGACCAACAGATTT 1980  
QY 1981 GTTCCAGCCCCGAAAGAAATCTGCGTTAGTATGCAATACCTATGAGGAGCTTTACT 2040  
Db 1981 GTTCCAGCCCCGAAAGAAATCTGCGTTAGTATGCAATACCTATGAGGAGCTTTACT 2040  
QY 2041 GACATTCGCTCTCTGACACACTGTTAGAGATCGGCGCACTGATGATGAAACACAAACAA 2100  
Db 2041 GACATTCGCTCTCTGACACACTGTTAGAGATCGGCGCACTGATGATGAAACACAAACAA 2100  
QY 2101 GATTCTGCGTTTCTCTCATGACGAATCTTCTGATAGACTGAGATGAAATGCGAAA 2160  
Db 2101 GATTCTGCGTTTCTCTCATGACGAATCTTCTGATAGACTGAGATGAAATGCGAAA 2160  
QY 2161 GGCCTCCGTCATACCTCTGAGAGCTACGTCATCGGTGGAAGTGTCTACACTCTTAAGAC 2220  
Db 2161 GGCCTCCGTCATACCTCTGAGAGCTACGTCATCGGTGGAAGTGTCTACACTCTTAAGAC 2220  
QY 2221 GACCTATTTACCTTTGCGTCTGCCATCTCTTTGCTAGAGCAAAAGATTTTATGCGT 2280  
Db 2221 GACCTATTTACCTTTGCGTCTGCCATCTCTTTGCTAGAGCAAAAGATTTTATGCGT 2280  
QY 2281 CACAACACCTAGAACCTTACGCTGAGAACCTTATTTCTCAAGCACTCATACCTACAA 2340  
Db 2281 CACAACACCTAGAACCTTACGCTGAGAACCTTATTTCTCAAGCACTCATACCTACAA 2340  
QY 2341 CCCCCAACTATTTAGATAGAGAACGAAAGTTTCTGAATCAGCTATAGAAAATTC 2400  
Db 2341 CCCCCAACTATTTAGATAGAGAACGAAAGTTTCTGAATCAGCTATAGAAAATTC 2400  
QY 2401 CCTAGGAAATTTCCCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 2460  
Db 2401 CCTAGGAAATTTCCCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 2460  
QY 2461 ATGGAAGGCACTATACCTCATATTCGACAAATCCGAAGGTTCTTGAGCAACAGATGATA 2520  
Db 2461 ATGGAAGGCACTATACCTCATATTCGACAAATCCGAAGGTTCTTGAGCAACAGATGATA 2520  
QY 2521 GCTGCGGATACGCGCTAGACCTTCTTTCTTCCAAACCAACATCTCTTTCAAG 2580  
Db 2521 GCTGCGGATACGCGCTAGACCTTCTTTCTTCCAAACCAACATCTCTTTCAAG 2580  
QY 2581 ACCCTTCAATCCAGAGATGAAGTGAATGTTATGATACAAATATGCTCTGAAA 2640  
Db 2581 ACCCTTCAATCCAGAGATGAAGTGAATGTTATGATACAAATATGCTCTGAAA 2640  
QY 2641 AGCTCTAGATGAGCGCGTGTATGATTTGATGGAAGGCTTAACTCTGATTCGTGTG 2700  
Db 2641 AGCTCTAGATGAGCGCGTGTATGATTTGATGGAAGGCTTAACTCTGATTCGTGTG 2700  
QY 2701 GGTGCGAAATTCGTGACAGGCGGATATCGGAAATTCCTACACCTATGATCTCAGATTC 2760  
Db 2701 GGTGCGAAATTCGTGACAGGCGGATATCGGAAATTCCTACACCTATGATCTCAGATTC 2760  
QY 2761 TTTGTTTCCGATGCTATCTGATGATGATGATGATGATGATGATGATGATGATGATG 2820  
Db 2761 TTTGTTTCCGATGCTATCTGATGATGATGATGATGATGATGATGATGATGATGATG 2820  
QY 2821 GACTCTTGGAATTTCCGCGTGGCAATCTTCAAGACAGGATTTTATCTAGAGGATGAC 2880  
Db 2821 GACTCTTGGAATTTCCGCGTGGCAATCTTCAAGACAGGATTTTATCTAGAGGATGAC 2880  
QY 2881 AACAACTACGCTACACCTCCAAATTTGAGCTCTTGGACATTAAGCTATGAAATCCGCT 2940  
Db 2881 AACAACTACGCTACACCTCCAAATTTGAGCTCTTGGACATTAAGCTATGAAATCCGCT 2940  
QY 2941 GGATCTTCAAGGAACCTACAAATGATGATGATGATGATGATGATGATGATGATGATG 3000  
Db 2941 GGATCTTCAAGGAACCTACAAATGATGATGATGATGATGATGATGATGATGATGATG 3000  
QY 3001 ACTCCGATGTTCTTCTAGGAGGTTTCTCATACTTTTAAAGGAATATTTGCTATAGGAA 3060



9	551.2	17.2	2793	9	AB1841	Sequence 15
10	498.8	15.6	16448	1	AE001587	AE001587 Chlamydia
11	498.8	15.6	26520	2	CPN133035	AJ133035 Chlamydia
12	498.4	15.6	2757	9	AB1837	AJ1837 Sequence 11
13	498.2	15.6	4926	2	CPU72499	U72499 Chlamydomoni
14	479.2	15.0	12127	1	AE002235	AE002235 Chlamydom
15	474.8	14.8	299650	2	AP002245	AP002245 Chlamydom
16	465.8	14.6	2787	9	AB1835	AB1835 Sequence 9
17	457.6	14.3	10757	1	AE001628	AE001628 Chlamydia
18	457.6	14.3	12676	1	AE002192	AE002192 Chlamydom
19	457.6	14.3	17280	2	CPN133034	AJ133034 Chlamydia
20	457.6	14.3	300650	2	AP002246	AP002246 Chlamydom
21	456	14.2	2815	9	AB1829	AB1829 Sequence 3
22	432.4	13.5	3000	9	AB1851	AB1851 Sequence 25
23	432.4	13.5	6030	2	CPOMP24	AJ001311 Chlamydia
24	423.2	13.2	2787	9	AB1839	AB1839 Sequence 13
25	341	10.7	2526	9	AB1833	AB1833 Sequence 7
26	361	10.7	10574	1	AE001586	AE001586 Chlamydia
27	307.6	9.6	2838	9	AB1849	AB1849 Sequence 23
28	300	9.4	3000	9	AB1851	AB1851 Sequence 25
29	298.2	9.3	6110	2	CPU65542	U65542 Chlamydomoni
30	263.8	8.2	6234	2	CPU65543	U65543 Chlamydomoni
31	256.6	8.0	10044	1	AE002237	AE002237 Chlamydom
32	255.8	8.0	10766	1	AE001585	AE001585 Chlamydia
33	252.6	7.9	3052	9	AB1831	AB1831 Sequence 5
34	191.8	6.0	840	9	AB1843	AB1843 Sequence 19
35	184.6	5.8	1830	9	AB1855	AB1855 Sequence 27
36	145.4	4.5	15789	1	AE002191	AE002191 Chlamydom
37	143.8	4.5	10695	1	AE001629	AE001629 Chlamydia
38	101.4	3.2	1200	9	AB1853	AB1853 Sequence 27
39	84.4	2.6	14482	1	AE002293	AE002293 Chlamydia
40	84	2.6	455	2	CPU59307	U59307 Chlamydomoni
41	74.8	2.3	14177	1	AE001360	AE001360 Chlamydia
42	62	1.9	216	2	CPU59306	U59306 Chlamydomoni
43	58.2	1.8	1545	9	AB1845	AB1845 Sequence 19
44	54.6	1.7	10001	1	AE002248	AE002248 Chlamydom
45	54.6	1.7	11781	1	AE001676	AE001676 Chlamydia

## ALIGNMENTS

AE01841 Sequence 15  
AE001587 Chlamydia  
AJ133035 Chlamydia  
AE18337 Sequence 11  
U72499 Chlamydophila  
AE002235 Chlamydo  
AP002545 Chlamydo  
AE001628 Chlamydo  
AJ133034 Chlamydia  
AP002546 Chlamydo  
AE1829 Sequence 3  
AE1851 Sequence 25  
AJ001311 Chlamydia  
AE1839 Sequence 13  
AE1833 Sequence 7  
AE001586 Chlamydia  
AE1849 Sequence 25  
AE1851 Sequence 23  
U65942 Chlamydophila  
U65943 Chlamydophila  
AE002237 Chlamydo  
AE001585 Chlamydia  
AE1831 Sequence 5  
AE1843 Sequence 17  
AE1855 Sequence 29  
AE001291 Chlamydo  
AE001629 Chlamydia  
AE1853 Sequence 27  
AE002293 Chlamydia  
U59307 Chlamydophila  
AE001360 Chlamydia  
U59306 Chlamydophila  
AE1845 Sequence 19  
AE002248 Chlamydo  
AE001676 Chlamydia

```

RESULT      1
LOCUS       A81827                      3200 bp    DNA
DEFINITION  Sequence 1 from Patent WO9858953.
ACCESSION   A81827
VERSION     A81827.1
KEYWORDS    GI:6791863
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 3200)
AUTHORS     Madsen, A. and Birkelund, S.
TITLE       NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL     Patent: WO 9858953-A 1 30-DEC-1998;
            MADSEN ANNA SORTE (DK); BIRKELUND SVEND (DK)
FEATURES
     source          location/Qualifiers
             1..3200
             /organism="unidentified"
             /db_xref="taxon:32644"
             205..>2987
             /note="unnamed protein product"
             /codon_start=1
             /protein_id="CAB69199.1"
             /db_xref="GI:6731864"
             /translation="MRTSIPWVLVSSVLAFSCHLOSLANELLSPDSSFNGNIDSGTFF
             IPTTSATYSLICDVFYEPKGTPLSDSCFKQTTDNILTEFGNHSLSLFGIDAGTHA
             GAASTTANKNLTFEFGSILSPDSPESTVYTGQGLTFSAGVLENIKRLVVGNGFS
             TADGAIKGSFLITGSGDALFNNSSSTRGGA IATTAGRIANNMTGYVRELNSIAS
             TSGCAIDECTSLSNKPLFYEGNNAKTTGCALCNTRKAGSPPELLISNKTLLFPAS
             VARTSGCAIHAUKLALSSGGFTFLRNWVSAPDKGAISIDASGEISLSELTNITTF
             VRNTLTGTSQDTPKRAINLIGSNGKFTFLRAANKHIIFFYDPLTSGTSSDYKINN

```

BASE COUNT				946 a	708 c	623 g	923 t
ORIGIN							
Query Match				100.0%;	Score 3200;	DB 9;	Length 3200;
Best Local Similarity				100.0%;	Pred. No. 0;		
Matches 3200;				Conservative 0;	Mismatches 0;	Indels 0;	Gaps
QY	1	CAATGTCGAAGAGACACTAACCCAGGAAATTCGATTTTCATTAACCCACTTTATTTA	60	GSAAGNPGCCTTGGCTTGGCTTTAAELKLVANLKSFPQVPSISGGKLLKQKVT FSOAGSLGMDSGTLLSTTAGSITTTNNGLKINVDLSGLKQPSLTKAKGSKNYK LNLIDLEGNLIESHMFSDQLFESLKITVDADVDIVNDLSILPYPAEDPNSBE OMVNWTTDPATNTKEATATATGTPGVSPERKSGALVCTMLMVFEDIDISLQO ATGNEHRQGFVWSMTNLFHKITDENKGRSFGVYIGVSGAHPEKDLFFRFD FARDKDFEIAHNNSRTYGGTLFPKHSHTLOPOLYULGKAPESLSIEKPEPE DVQVFSHSDNRMEHTHTSLPEEGSKSNEDLNGIGLIDLPYLSNPHLFFKTC KVEKVYVSQNSFEPSSDSGRFSIGRLNLISIPVGNFVQGDIGDSYTDLDSG VYRNRPSTATLVMSIPDSMKIRGNLSRQAFLLRGSNYVYNSNCELFGHYTAM SRNWNVDVGTCLR"			
Db	1	CAATGTCGAAGAGACACTAACCCAGGAAATTCGATTTTCATTAACCCACTTTATTTA	60				
QY	61	AATTCCTTCGCGCATATATAAATAGAAAACAGAGAGCAAGATAAAAATCTTGAC	120				
Db	61	AATTCCTTCGCGCATATATAAATAGAAAACAGAGAGCAAGATAAAAATCTTGAC	120				
QY	121	AGCTGTTTTGTCATCTTTAACTGATTTACTATTTTGTTCATATTTGATGCGAATAGT	180				
Db	121	AGCTGTTTTGTCATCTTTAACTGATTTACTATTTTGTTCATATTTGATGCGAATAGT	180				
QY	181	TCTCTAAAAAAACAAAGACATTACCATGAACCTTCGATTCCTTG6GTTTAGTTTCCTCC	240				
Db	181	TCTCTAAAAAAACAAAGACATTACCATGAACCTTCGATTCCTTG6GTTTAGTTTCCTCC	240				
QY	241	GTGTTAGCTTCCTCATGCACTCACTCAAGCACTACCTACCTAAGAGAACTTTATCACTGAT	300				
Db	241	GTGTTAGCTTCCTCATGCACTCACTCAAGCACTACCTACCTAAGAGAACTTTATCACTGAT	300				
QY	301	GATAGCTTTAATGGAATATGCATTCAGGAACGTTTACTCTCCAAAACCTTGACGCCACACA	360				
Db	301	GATAGCTTTAATGGAATATGCATTCAGGAACGTTTACTCTCCAAAACCTTGACGCCACACA	360				
QY	361	TATTCCTCAACAGAGATGCTCTTTTACGAGCCTGGAAAAGCACTCCCTTATCTGAC	420				
Db	361	TATTCCTCAACAGAGATGCTCTTTTACGAGCCTGGAAAAGCACTCCCTTATCTGAC	420				
QY	421	AGTTGTTTTAGCAAAACACAGGAACTTACCTCTTG6GGAAGGTCATAGCTTAAAG	480				
Db	421	AGTTGTTTTAGCAAAACACAGGAACTTACCTCTTG6GGAAGGTCATAGCTTAAAG	480				
QY	481	TTTG6CTTTAATAGATGCTGGGACATATGCAAGTCTCTGATCTTACACAGCAAAATAG	540				
Db	481	TTTG6CTTTAATAGATGCTGGGACATATGCAAGTCTCTGATCTTACACAGCAAAATAG	540				
QY	541	AATCTTACCTTCTCAAGGTTTTCCTTACTAGATTGATTCCTCTGACACAAAGGTT	600				
Db	541	AATCTTACCTTCTCAAGGTTTTCCTTACTAGATTGATTCCTCTGACACAAAGGTT	600				
QY	601	ACTACAGGTCAGGGAGCGCTTCTCAGACAGAGAGCGTAAATTTAGAAAATATTCGFAAA	660				
Db	601	ACTACAGGTCAGGGAGCGCTTCTCAGACAGAGAGCGTAAATTTAGAAAATATTCGFAAA	660				
QY	661	CTTGTAGTTGCTGGGAAATTTTCTACTGCAAGATGNTGGAGCTATCAAGAGAGGCTTTTC	720				
Db	661	CTTGTAGTTGCTGGGAAATTTTCTACTGCAAGATGNTGGAGCTATCAAGAGAGGCTTTTC	720				
QY	721	CTTTTAACTGGCACTTCGAGATGCTCTTTTATAGTAAACACTTTCATCAACAAAGGA	780				
Db	721	CTTTTAACTGGCACTTCGAGATGCTCTTTTATAGTAAACAACCTTTCATCAACAAAGGA	780				
QY	781	GGACCAATTCCTACTACAGACGGCGCTGCATATGCAAAATACACAGGTTATGTTAGATTG	840				
Db	781	GGACCAATTCCTACTACAGACGGCGCTGCATATGCAAAATACACAGGTTATGTTAGATTG	840				





QY	1618	GGGAGTFTTACATCAGCAACCTAGGAATCAATGTTGACTCCTTAGGGCTTAAGACGCC	1677
Db	1544	gaggaggtcactttaaaaggtcttccacttccgttagtaccttctttagcgaggttaagaa	1603
QY	1678	GTGCGCCTAACAGCAAAAAGTGCTCTCAATTAAGATGATCGTATCTGGGAAGCTCAACCTG	1737
Db	1604	gttgaattgtcgtctctctgcgcgaagtaaaatgtacgaccttagtgcgtaccttctt	1663
QY	1738	ATTGATATTGAAGGGAACATTATTGAAAGTCATATGTTAGCCATGACCACTCTTCTCT	1797
Db	1664	ttagataccaaggaatgtctatgataaalcacgaacttaggaaaaaactcaagacttttca	1733
QY	1798	CTATTAAAAATCAGCGGTGATGCTGATGTGTGATTAACGTTTACATCAGACAGCGCTTATC	1857
Db	1772	gcggttctctcagtagcaactcctcagcactatggtatcaagtaacttgggaatgact	1831
QY	1918	TGGACTCTACGATACAG-----CTACAATATCAAAAAGGCGACGGCACTTTGGACCAA	1971
Db	1832	tgggtttgatatcagcgcaagcaactccaagaactcaagacagacattagctttagccaat	1891
QY	1972	ACAAGATTTTCTTCCCAAGCCCGCAAAAATAATCTGGGTATGTATGCAATAACCTTATGGGA	2031
Db	1892	acaggtcactctcgaatcctgactgtagcgtagaagacctttagtctcctaaatagccttggga	1951
QY	2032	GTCCTTACTACATTGGCTCTGTCACACAGCTTTGAGAGATCGCGCAACTGATATGGAA	2091
Db	1952	tcttttccagacatccacgaagcttcaatggttcataagagaagtgctttagcttctgt	2011
QY	2092	CACAAACAAGGTTCTTGCGGTTCTTCCATGACAGAACCTTCCATTAACATCGAGAGATCA	2151
Db	2012	tcgactcgaggtctctgggtcgtcgaggagtcgcgaacttcttagataaaagataaagaaggg	2071
QY	2152	AATCGCAAGGCTTCGCTCATACCTTGGAGGCTACGTCATCGGTGGAAGTGTACACT	2211
Db	2072	gaaaaaagcacaatcacgcatataaacttgyttagatctgcatgtagaggtgcagcgcaact	2131
QY	2212	CTTAAAGACACCTATTATACCTTGGGCTTCCGCATCTCTTGGCAGAGACAAGATATGT	2271
Db	2132	tgcttgcgaacttaattagctttagcttcttgccaactccttggtagcgataaagaatttc	2191
QY	2272	TTTATGCGCTACACAACTCTTAGAACCTACGAGTGAAGCTTATTTCTTCAGCACTCAT	2331
Db	2192	ttagctgcataaactatcactgataccttagtagcagaagcctctatataccaac-----	2245
QY	2332	ACCCTACCAACCCCAAAACTATTATTGACATTAGAGAAAGCAAAATTTTTCGATACGCTATA	2391
Db	2246	attacgaatagttagtgggtctatagtttgccttcttagataaaactccggcctcttggagt	2305
QY	2392	GAAAAATTTCCCTAGGAAAAATTCCTCCATGCCCTTGGATGTCACAATTCTGTTACGCATTC	2451
Db	2306	cataaa-----ccctcgttttagaagggcagctcgctatagccaagtc	2350
QY	2452	GACACACGCTTAGAAGACGACATTAACCTATTTGCCAGAAATCCGAAGGTTCTTGAGCAAC	2511
Db	2351	agtaatgattctgaagacaagatactactgctatcctcgagtgtagaaggttcttgggggaat	2410
QY	2512	GAGGTGTACTGTGTGTATGCGGCTAGACCTTCCCTTTTGTTGCTTTCACACCCACATGCT	2571
Db	2411	aaigtctttaacaatgtagtggaggtcttcttcaactcttatacccgaaataactgcatgt	2470
QY	2572	CTTTTCAAGACCTTCAATCCACAGATGAAGATCGAATGTTTATGTATCAACAAAATAGC	2631
Db	2471	ttt-----gataccttatgctcatataccaacatcgaaatctgacctatatacgcagagacgc	2527
QY	2632	TTCTTCGAAAGCTCTATGTATGATGGCCGTGGTGTATTATTATTTGGAAGCGTCTTAACCTCG	2691
Db	2528	ttctcggagagaaggtcagaagaagaatccttlttagtagacagaacaccttcaatttact	2587

OY	2692	ATTCCTGGGTGGAAATTCGCGACAGGGGGAATATCGAGAAATTCACACATATATTC	2751
Db	2588	ttgcctatagggatgaaagtttgggaagttctctgattgtcaatgaactttcttatagctg	2647
OY	2752	TCAGGATTCCTTGTGTTCCGATGTCATACGTAACAAATCCCAATCTACAGGACTTGTG	2811
Db	2648	acttatactatgttctctgattctatctcggaaatgataccaatgactacagcattgta	2707
OY	2812	ATGAGCCGACAGCTTCGTGAAATTCGGGGGCGCAATCTTTCAGACAGGACTTTTACG	2874
Db	2708	atcagcggaagcctcttgggaaacttatgccaataacttgcacgaacagcccttgcagt	2765
OY	2872	AGGGGTGACCAACAACGTCTACAACCTCCAAATGTGAGCTCTTGAGACATTACGCTATG	2931
Db	2768	ctgtcagcagcagctacactacgcctctctcctatgtttgaaagtgtcgcgcagtttgtctt	2822
OY	2932	GAATCGGTGGATTCCTCAAGAACTCAATGATGATGTTGATGACCAAACTCGGATTCAG	2991
Db	2828	gaagtttcggatgattcccaagatttaaaagtatgactcttgggggaattgaattccaattcag	2887

Search completed: October 2, 2001, 03:56:46  
Job time: 52753 sec



Db 967 gcaattctatactgagcagctgagagtgtagtcttcaagcaagcagggacattacc 1026  
 QY 1204 TTTGTAGAAATACCTTTCACCAACCGGAGTACCGATACCTTAAGCTAATGGGATC 1263  
 Db 1027 ttcacatggaaatgcaatctgtctgcaactacac--gacaaactacaanaagaattatct 1083  
 QY 1264 AACATTTGAGGAGTACGGGAAATTCACGAAATTCAGGGCTGTAAAAATCATCAATTTTC 1323  
 Db 1084 gacataagctactacgcgaagaatcaagattatcagtgcaatatctggagcatatgcattct 1143  
 QY 1324 TTTATATATCCATCAC-----TTCAGAGGAACCTCATCAGACGTATATGAGATTAAT 1377  
 Db 1144 tctcagactcagattactctactaataagctgctgagattctcagactatctaattcact 1203  
 QY 1378 AACGCTCTGGCGGAGCTCTCATATCATATCAAGGACATTTCTATTTCTTGAGAAACC 1437  
 Db 1204 aaggtctgagcgggtaataagtaagatagtagtgagtgagtgatttcttctgtggaag 1263  
 QY 1438 CTACACGACAGATGATTAAGTTGCTGACAAATTTAAATCTTCATTCACGACCGACATC 1497  
 Db 1264 ctctcgaagatgaaagcaaaagtctgacgacaactcactctcagctgagacgacctgt 1323  
 QY 1498 TCCCTATCCGAGGAAATTTATTTGCTACAAAAGGAGCTCTTTAGAGACAGAGCTTC 1557  
 Db 1324 actctaacgcaggaatttagtacttaaacgtgtgtcactcctcgatagcaagagctt 1383  
 QY 1558 TCTCAAGAGCGCGTCTCTCCGCGCATGAGATTCAGAGCAACCATATCACTACAGCT 1617  
 Db 1384 actcagacccggggtctctctgtattatagtagtgagcgacaaacgtlaaagaactaa 1443  
 QY 1618 GGGAGATTTATGATACAGCAACCTTAGAATCAATGTTGACTCTTAGGTTTAAAGCCGC 1677  
 Db 1444 gaggagatcacttaacagcttctcactctgtagactctttagcgaggtagaaga 1503  
 QY 1678 GTACAGCTTACAGCAAAAGTGTCTCAATAAAGTATGATCTGTGGAGACTCAACCTG 1737  
 Db 1504 gtgttaattgcctctctgtcagcaaglaaagtagccttagtgtgtcagcttctct 1563  
 QY 1738 ATTGATTTATGAGGAAATTTATGAAAGTCAATATGTTACGCATACAGCTCTCTCTCT 1797  
 Db 1564 ttggataaccaaaggaaatgtctatgaaatacaagacttgagaaatacaagaacttca 1623  
 QY 1798 CATTTAAAAATCAGGGTTGATGCTGATGTTGATGATGATGATGATGATGATGATG 1857  
 Db 1624 ttgtgcagctctctgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1671  
 QY 1858 CCTGTTCCCTGAGAGATCTTAATTAAGATAGAGATTCAGAGCAATGATGTTAT 1917  
 Db 1672 gcggttctctacagtagcaactctcactagcactatggtatcaaggtacttggggaatgact 1731  
 QY 1918 TGGACTACGATACAG-----CTACAAATACAAAAGAGCGCACGCACTTGACCAAA 1971  
 Db 1732 tgggttgatgataacggaagcaactcacaagaactaagaacatactagctgtgacacat 1791  
 QY 1972 ACAGATTTGTTCCAGCCCGCAAGAAATCTGCGTTAGTATGCAATACCTATGGGGA 2031  
 Db 1792 acaaggtacctctcgaatcctgagcgtaagaacctttagtcttaataagccttgggga 1851  
 QY 2032 GTCCTTACTGAGATTCGCTCTGCAACAGCTTTGAGATATCGGCCAATGCTATAGGA 2091  
 Db 1852 tcttcttgaacatccaagcgaatccaagtgtaataagtagaagaagtgctttagactctgt 1911  
 QY 2092 CACAAACAGAGTTTCTGGTTCTCTCCATGACGAATCTCTCATTAAGACTGGAGATGA 2151  
 Db 1912 tcaatcagaggtctctgtggtctgcggaagtcgcaaatcttcttagataaagaagaag 1971  
 QY 2152 AATTCGCAAAAGCTTCGCTCTGATACCTTGAGAGCTTACGTCAGTGGTGAAGTCTCACT 2211  
 Db 1972 gaaaaacgcaaatcacgtcataatctgtgtgatatcatcggaagtgcaagcaact 2031  
 QY 2212 CCTAAAGAGACCTTATTTACCTTTGGTTCTGCAATCTTTGCTGAGACAAAGATTTCT 2271  
 Db 2032 tgtctgaaaaactaattagcttgccttctgcaactcttctgttagcgataaagaattc 2091

QY 2272 TTTATCGCTCACAAACAATCTGAACCTACGGGAGAACTTTATTTCTCAAGCACTCAT 2331  
 Db 2092 ttagtcgctaaaaatacatctatctatctatcgagagccttctataccaac----- 2145  
 QY 2332 ACCCTACAAACCCCAAAACTATTTGAGATTAGGAAGACAAAGTTTCTCAATCAGCTATA 2391  
 Db 2146 attacagaaatgtagtgggtcatagtgctgtccttagataaacttcctgtgccttggagt 2205  
 QY 2392 GAAAAATTTCCCTAGGAAATTCCTCAGCCTTGAGATGTCGAATTTGCTTCAAGCTTCA 2451  
 Db 2206 cataaa-----ccctcgtttagaagggcagctcgttataagccagctc 2250  
 QY 2452 GACACCGTATGGAAGACGACATACCTATTCATTCGCCAGATTCGGAAGCTTCTGAGCAAC 2511  
 Db 2251 agtaagatctgaagaacaaagatatactgcgtatcccgaggtgaagaagctctgggggaat 2310  
 QY 2512 GAGTGATATGCTGTGTGTATCGGCTAGACCTTCTCTTTGTTCTTTCACACCAATCTCT 2571  
 Db 2311 aatgctttaaactatgtagtggagcttctctctcattcttactatccatgaatcactgactgt 2370  
 QY 2572 CTTTTCAGACCTTCTATTCACAGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2631  
 Db 2371 ttt--gatcctatgctccatatacatcaactgaactgaactataatagtagcagcagc 2427  
 QY 2632 TTTCTGGAAGCTCTATGATGCGCGGTTTATGATGGAAGGCTGCTTAACTCTCTG 2691  
 Db 2428 tctcggagaaaggtacagaagaagatctttagatgacagcaactcttcaattatct 2487  
 QY 2692 ATTCTGTGGTGTGCAAAATTCGTGACAGGGGGAATTCGAGATTCCTACACCTATGATCTC 2751  
 Db 2488 ttgcctataggggttgaagtttgaagttctctcgtatgtaagacttcttactatgactc 2547  
 QY 2752 TCAGATTTCTTTGTTCCGATGCTATTCGTACCAATCCCAATCTACAGCACTCTGTG 2811  
 Db 2548 acttatacctatgctcgtatctatcgcgaatgacccaatgcaactgaacgaactgt 2607  
 QY 2812 ATGAGCCGACGACTCTTGAAGAAATTCGCGGTGCAATCTTCAAGACAGCAATTTTATG 2871  
 Db 2608 atcagagagcctcttggaataactatgccaataacttagcagacagagccttgcaagt 2667  
 QY 2872 AAGGCTAGCAACACTAGCTGTACCACTCCAAATTTGAGCTCTTCGCAATTAAGCTATG 2931  
 Db 2668 cgtgcagcagcactacagcctctcctcctatgattgaaggtgtcgcgcaggttgcctt 2727  
 QY 2932 GAACCTCCGCGATCTTCAAGAACTCAATGTAGAGTGTGATGATGATGATGATGATGATG 2991  
 Db 2728 gaagtcgtgagctcctcaggaattataatgtagatcttgggggtgaagttccaattcag 2787  
  
 RESULT 15  
 AAA30849  
 ID AAA30849 standard; DNA; 2950 BP.  
 XX  
 AC AAA30849;  
 XX  
 DT 29-AUG-2000 (first entry)  
 XX  
 DE Chlamydia antigen CPN100635 full length coding sequence.  
 XX  
 KW Chlamydia antigen; diagnosis; infection; community acquired pneumonia;  
 KW therapy; upper respiratory tract disease; bronchitis; sinusitis;  
 KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;  
 ds.  
 OS Chlamydia pneumoniae.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 101..2887  
 FT /tag= a  
 FT /product= Chlamydia antigen CPN100635  
 FT sig\_peptide 101..229  
 FT /tag= b

Db 100805 cggagcctctgggaactatgcaactagcagcagcagccttgcaagtcgctgc 100864  
 QY 2877 TAGCAACAACACTAGCTGTACACACTCCATGTGAGCTTTGGGACATTCAGTATGGAAC 2936  
 Db 100865 aggcagcactacacgctctctcctatgctgaagtcgcgcacgcttcttctgaagt 100924  
 QY 2937 CCGGATCTTCAAGAACATGTAGATGTGTACCAAACTCCGATTTCTAG 2991  
 Db 100925 tcgtgacctcctcaagcagattataatgtagatcttgagggtacgtaagttccattcag 100979  
 RESULT 14  
 ID AAX06817  
 XX AAX06817 standard; DNA: 2815 BP.  
 AC AAX06817;  
 XX  
 DT 26-APR-1999 (first entry)  
 XX  
 DE Chlamydia pneumoniae surface exposed protein Omp5 DNA.  
 XX  
 KW Omp5, outer membrane protein 5; surface exposed protein; antigen;  
 KW Infection; diagnosis; vaccine; atherosclerosis; asthma; ss.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..2787  
 FT /tag- a  
 PN MO9858953-A2.  
 PD 30-DEC-1998.  
 XX  
 PF 19-JUN-1998; 98MO-DK00266.  
 PR 23-JUN-1997; 97DK-0000744.  
 XX  
 PA (BIRK/) BIRKELUND S.  
 PA (CHRI/) CHRISTIANSEN G.  
 XX  
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;  
 PI Mygard P;  
 DR MPI: 1999-105610/09.  
 DR P-PSDB; AAM88418.  
 XX  
 PT Species-specific test for identifying mammals infected with  
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for  
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding  
 PT these proteins  
 XX  
 PS Claim 6; Page 42-43; 115pp; English.  
 XX  
 CC This DNA sequence codes for the novel 97.2 kDa surface exposed  
 CC protein Omp5 (see AAM88418) of the human respiratory pathogen  
 CC Chlamydia pneumoniae. By generating antibodies against C.  
 CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)  
 CC was obtained which reacted with outer membrane proteins. The  
 CC antibody was used to identify the genes (see AAX06816-27) encoding  
 CC Omp5-Omp5 proteins (see AAM88417-28) in an expression library of  
 CC C. pneumoniae DNA. The genes are situated in 2-gene clusters:  
 CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in  
 CC the other, and encode polypeptides of about 89.6-100.3 kDa and  
 CC about 56.1 kDa. The invention provides a new species specific test  
 CC for identifying mammals (including humans) infected with Chlamydia  
 CC pneumoniae. The test comprises detecting antibodies specific for  
 CC Omp5 or detecting nucleic acid fragments encoding these outer  
 CC membrane proteins, especially by PCR. The proteins are also used  
 CC in the diagnosis of C. pneumoniae infection in mammals. The  
 CC nucleic acids and proteins can also be used in the immunization of  
 CC mammals, the nucleic acids being particularly useful as DNA

CC vaccines for effecting in vivo expression of antigens. The  
 CC vaccines may also prevent atherosclerosis and bronchial asthma,  
 CC which are possibly associated with C. pneumoniae.  
 XX  
 SQ Sequence 2815 BP; 790 A; 575 C; 626 G; 824 T; 0 other;  
 Query Match 14.2%; Score 456; DB 20; Length 2815;  
 Best Local Similarity 51.8%; Pred. No. 6,2e-105;  
 Matches 1430; Conservative 0; Mismatches 1240; Indels 90; Gaps 14;  
 QY 274 GCTAACGAGAACCTTTATACCTGATGATAGCTTTAATGAAATATCGATTGACGACG 333  
 Db 76 gcaactgctgaaatatagccctctgtagcttgtagcgaagtactaacacagcgacc 135  
 QY 334 TTTACTCCAAAACCTTAGCCACACACA-----TATCTCTAACGAGAGATGCTTTT 387  
 Db 136 tatactcctaaataaatacgaactggaatagactatactgcagaagataataactcg 195  
 QY 388 TAGAGCCTGGAAAAGGACGACCTCCCTATCTGACAGTGTGTTTAAGCAAAACGAGACAT 447  
 Db 196 caaaaccttgaggatcgcgagctttaagaaaggtgttctcgaactacgaatct 255  
 QY 448 CTAACTCTTGGGGAACGGTCAATAGCTTAAGCTTGAAGTGGCTTGAAGTGGCACTCAT 507  
 Db 256 ttaagcttgcgtaagggtactcaactcttctttaaataaagctct---agtgct 312  
 QY 508 GCAGTCTGCTGCATCTTACACAGCAATTAAGAAATTTACCTTTCAGAGTTTCTT 567  
 Db 313 gaagcgcaagcacttctgttaacaactgaataaactgtcgtcaaaagatttcgagt 372  
 QY 568 CTGAGTTT-----TGATTCCTCTCTAGCAACAGGTTACTACAGTTCAGGTCGAGACGCT 621  
 Db 373 cttaacttcttagcgcgcccatatcggtatacaaacccctcaggaagaagtgcaagt 432  
 QY 622 TCCTCACAGAGAGCGCTTAATTTAGAAAATATTCGTAACCTGTAGTTGCTGGCAATTT 681  
 Db 433 aaatgtagagggtactctacattgtatacaatcaggaacttattatlaaacaagattac 492  
 QY 682 TCTACTGCACATGCTGAGCTATCAAGAGCGCTTCTTTAACTGACACTTTCGGA 741  
 Db 493 tgtgaggaanaatgcygagcacttctaccaagaatcttcttgaaaaacaacagcgga 552  
 QY 742 GATGCTCTTT-----TAGTAACAACTCTTCATCAACAAGGAGAGCAATTCGT 792  
 Db 553 tcgatttcttcttaaggaataatcgagcgcaacaggaanaagtggtgctatttgt 612  
 QY 793 ACTACAGCAGCGCTCGCATAGCAATTAACACAGGTTATGTAGATTCTTACTAACA 852  
 Db 613 gctactgtagctgtagatattacaataatagcgtctcctcctctcgaacaatatt 672  
 QY 853 GCGCTTACGTCAGAGCGCTATCGATGATGAAGGACGCTGATCTATGAAACAACAA 912  
 Db 673 gctgaagctgtagtgtagctataaataagcacaggaactgtaacattacaaggatagc 732  
 QY 913 TTTCTAATTTTGAAGGAATGCAGGAAATACATGCGCGTGCATCTGCAACACCAAG 972  
 Db 733 tccctgtattctcgaataatagtgacagcgacgcaaggaatggaagactctt-- 790  
 QY 973 GCGAGTGATCTCCTGAACATGATATCTTAACATTAAGACTGTGATCTTGTCAAA 1032  
 Db 791 -----cggagatgcgtagtctaccatatctgggaatcagaagttaacttctcagaagaac 846  
 QY 1033 GTAGCAAAACAAGCGGTGGCGCATTCATGTAATAAAGCTAGCCCTTTCCTC- -T 1086  
 Db 847 caagctgtagcgaatgcygagcacttattagtaagaagttacactggtctcgggggg 906  
 QY 1087 GGAGGCTTTACAGATTTCTTACGAATTAATGTCTCATCAGCAACTGC- -TAAGGGGGGT 1143  
 Db 907 ggggggggtagtctccttcttcaacaatagctcgaagtaacccctgcaagtaagtgtga 966  
 QY 1144 GCTATCAGCATCATGCGTCCAGAGAGCTCAGTCTTTCGCAACAGAGCAACATTAAC 1203

QY	686	CTCAGATGGGAGGAGCTTCAAAAGAGGCTGTCCTTTCCTTTAACTGGCACTTCTGGAGATG	745
Db	98650	aggaanaatgagcgagccactctaccagaaactctctcttcttgaanaaagcagcagggatcga	98749
QY	746	CTCTTTT-----TAGTAACAACCTTTATATCAACAAGAGGAGACCAATTGGCTACTA	796
Db	98750	ttctctttagaaggaataaactcgagcgcaacagggaaaaaaggctggctatttcttgctga	98809
QY	797	CAGCAGCGGCTCGCATAGCAATAATACACAGCTTATGTTAGATTCTTATCTAACATAGCT	856
Db	98810	ctggtactcgtatgataatacaataataacggtccctaccctctcttcgcgaacaatatctgtc	98869
QY	857	CTACCTCGGAGGCGGCTTCATGATGATGAAGGACGCTCATCTATACGAACAACAATTTC	916
Db	98870	aagctgcagcgvgggagcctaaatagcaagaagaaactgcacaatttcagggaaatacgtctc	98929
QY	917	TATATTTTGAAGGAGATGCAGGAAAACACTACTGGCGGTGCAGTCTGCAACCAAGGCGA	976
Db	98930	ttgtattcttcgtaaaatagltgagcaagcgaccgcaggaatctgaag-----agctcttt	98983
QY	977	GTGGATTCCTCGAAGCTGTAATTCCTAACAMPDAGACCTGATCTTGCTTCAACAGTAG	1036
Db	98984	ctggagatgacgaatgltacacatactgggaatcagaatgtaactctctcaggaacaacag	99043
QY	1037	CAGAAACAAGCGGTGGGCCCATCTCATGCTTAATAAAAGCTAGCCCTTTCCTC-----TGAGAG	1091
Db	99044	ctgtagactaagcgagcgacacttaagtaagaagccttaacactgcttcgcggggggggg	99103
QY	1092	CTTTACAGAGTTTCTACGAATAATGTCTCATACGACACTCC---TAAGGGGGTGCTAT	1148
Db	99104	gggtactctctcttcttaacaataatagttcaagaagtaaccactcaagtaatgtyggagccat	99163
QY	1149	CAGCATCGATGCTCTCAGAGAGAGCTCAGTCTTTCGCAAGACAGGAAACATTACTCTTGT	1208
Db	99164	ttctatactgacgactggagagtgagcttcttcagcagaagcaggggacattactctcaa	99223
QY	1209	AAGAAATACCTTTACACACACCCGGAAGTACCGATACTCTTAACGTAATGCGATCAACAT	1268
Db	99224	tgggaatgccaatgtltgtaaac---taccaccaaaactcaaaaaaagaaatctctatgacat	99280
QY	1269	AGGAAGTAAACGGGAAATCTCAGGAAATTCGGGGCTCTTAAATACTTAAATTTTCTCTA	1328
Db	99281	aggaatctactgcaaaagatcaagaattctacgcaatctcggcaatctcggcctagacatcttctta	99340
QY	1329	TGATCCCATCACC-----TTCAGAAAGAACCTCATCAGACGATTTGAAGTAATTAACGG	1382
Db	99341	cgatccgattactcgtctaatacgcgtcgcgattctctacagatacttcaatctcaataaagc	99400
QY	1383	CTCTGGCGGAGCTCTCATATCATATMCAAGGAACGATTTCTATTTCCTGGAAACCCCTAAC	1442
Db	99401	tgaatgcaggaataagtaagaataatagtgggucgagtgttcttcttgtagaagaagcttc	99460
QY	1443	AGCAGATGCAACTTAAAGTGTGTGCAATTTAAATCTTCATTCAACGACGACAGTCTCCCT	1502
Db	99461	tgaagatggaagcaaaagtgtgcagcaaacctcaactctacgctctgaaagcagcctgtaacct	99520
QY	1503	ATCCGGAGGAAAGTTATTGCTACAAAAGGAGTCACTTTAAGAGACAGAGCTTCTCTCA	1562
Db	99521	aactgcaggaatattagtaacttaaacgtyggtcactctcgatacgaaaagccttactacca	99580
QY	1563	AGAGGCGGCTTCTCTCCGCGCATGAGATTTCAGGAACGACATTATCACTTAACAGCTGGGAG	1622
Db	99581	gaccgcggggtccctcgttatataatgagatggcggcacaaagcttlaaagaagaatacagaagga	99640
QY	1623	TATTACAATTCACGAACCTAGGAATCAATGTTGACTCTTAGGTCTTTAAGCAGCCCTCAG	1682
Db	99641	gttcaactttaaagctcttccatctccgttagactctttaagcgagggtaagaagatgct	99700
QY	1683	CCCTAAACGCAAAAGGTGCTTCAATTAAGTGATGTGGAAGACCTCAACCTGATTTGA	1742
Db	99701	aattgctgcttcctgcaggaagtaaaaatgtagcccttaagtccttaagtcctctctcttggga	99760

OY	1743	TATGAAGGGAACATTATATGAATCATCATATGTTCCAGCCATGACCAAGCTCTTCTCTATTT	1802
Db	99761	taaccaagggaaagctcatactgaaataacgcgactctgaagaaataccaagaacttccatttgc	99820
OY	1803	AAAAATCACGGTTATATGTCATGATGTTTACTTAACAGTTTGACATCAGACGCTTATACCTCTG	1862
Db	99821	gcagctctctctctggttactgtaaacactcaagatg-----ttccagcgt	99868
OY	1863	TCTCTGTCGAGATCTTAATTACAGAAATACGGAATTCGAAGACATGCAATGTAATGTGAC	1922
Db	99869	tcttacaagtagcaactccctacgcactatggttcatcaagtagtctgggaatgacttgggt	99928
OY	1923	TACGATATCACG-----CTCAATATACAAAAGAGCCGACGCAACTGTGACCAAAACAG	1976
Db	99929	tgatgataccgaacagcaactccaaagactlaagaacagacatagcttgagaccataaag	99988
OY	1977	ATTGTGTCCAGCCCGGAAAGAAATATGCTGTATGATCAATACCTTATGGGAGTCTT	2036
Db	99989	ctaccttcggaatccctgagcgtcgaagaaccttgcttccataagaccttggggatcttc	100048
OY	2037	TACTGACATTCGCTCTCTGCACACGCTTGTAAGATCGCGCAACTGTAATGGAACACAA	2096
Db	100049	ttcagacatcccaagcgatctcaagtgctcatagagaagatgcttgctcttgcttcaga	100108
OY	2097	ACAAAGTTTCGGGTTCTCTCCATGACGAGAACTCTCTGCATTAAGACTGGAATGAAAAATCG	2156
Db	100109	tcgagtgcttcgggcgtcgcgagctcgcaacttcttagataaagataagaaggggaaaa	100168
OY	2157	CAAAAGCTCCGCTATACCTCTGAGCGTCAGTCATGAGTGGAAGTCTCACACTCTTAA	2216
Db	100169	acgcaataaccgctcataatcttggtagatatgctatcctcagggggtgcagcgcgaacttgc	100228
OY	2217	AGACGACCTATTTAACCCTTTCGCTTCGCCATCTCTTGTCTAGAGACAAAGATTGTTTAT	2276
Db	100229	tgaaaacttaactagcttggcttgcactcttggtagcgataaagatctcttagt	100288
OY	2277	CGCTCACACAAACCTGTAACCTGAGGTGGAACTTTATCTCAAGCACTCATACCT	2336
Db	100289	cgcataaataatactgatactatgcagtagcgcttctatataccaacaactc-----	100342
OY	2337	ACAAACCCCAAAACATTTTGAATTAAGAAAGACAGCAAACTTTTCTGAACTACACTATAGAAA	2396
Db	100343	agaatctagtggttcatagtttgcctcttagataaactcttcggccttggagtcataa	100402
OY	2397	ATTCCTTAGGGGAAATTCCTCCCTAGCCTTGATGATGCCAAAGTTTCGTTACGCCATTCAGACAA	2456
Db	100403	a-----ccctcgctttagaagggcagctcgctatagcagcagtcagtaa	100447
OY	2457	CCGATATGGAACGCACATACCTCATTTGCCAGAAATCCGAAGGTTCTTGAGCAACAGATG	2516
Db	100448	tgactctgaagcaaaagataactgctgatactctgaggtgaaggttctcttggggataatgc	100507
OY	2517	TATAGCTGTGTGATTCGGGCTTAGACCTTCCTGTTTCCAAACCCATCCTCTCTTT	2576
Db	100508	ttttaaactgactgttggagagcttcttccacttctacactcgaataactcgaa---ttgctt	100564
OY	2577	CAAGACCTTCAATTCACAGATGAAGATCGAAATGGTTTATGTATCACAAAATAGCTTCTT	2636
Db	100565	tgataactatgtctcataataatcaaacgaatctgacctataatgctgaagaacagctcttc	100624
OY	2637	CGAAAGCTTAGTGAATGGCGGTGTTTATGATTAATGGAAGCTGCTTAACCTCTCGATTC	2696
Db	100625	ggagaaagtgacagaagaagatctctttagtagaagcaacctcttcaattatcttgc	100684
OY	2697	TGTGGGTGCGAAATTCGGACGGGGGATATCGGAGATTCCTACACCATATATCTCAGG	2756
Db	100685	tataggggtgaagtttgagaagttctccgatgtaaaagacttctcttaagctcgacttc	100744
OY	2757	ATTCCTTTTCCCATGCTCATCTCTTAACAATCCCAATCTACAGCGACTTGTGATGAG	2816
Db	100745	atcctaigtctctgactcttatacgcgaatgataccaatatgatacgaagcacttgaatcag	100804
OY	2817	CCCAAGACTCTTGGAATAATTCGGCGGTGGCAATCTTTCAAGCAGGCAATTTTACTGAGGG	2876

```
QY 2392 GAAAAATTCCTAGGAATTCCTCCCTTGAGTGTCCAAGTTGCTTCAGCCATCA 2451
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2298 -----ttcttaagaagatccctgtctctcttccaggaacccctagctaacccatagc 2350
QY 2452 GACAAACCGTATGAAACCCACTATACCTCATTTGCCAGAACGTTCTTGGAGCAAC 2511
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2351 gataagacatcgaaacccaagatacaatactactactgttaagaagctggggagat 2410
QY 2512 GAGGTATGATGCTGGATGGATGCTAGACCTTCTTTGTTCTTTCACCCACATCTC 2571
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2411 gatagtttcgctttagaatcttggtggaagagctccgatttgccttagatga---aagtgct 2467
QY 2572 CTTTTCAGACCTTCTTTCACACATGTAAGTCAAAATGTTATGATCAAAAATAGC 2631
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2468 ctatttgagcagctacatgccttcaatgaattgcaatttgccttgcacatcaggaaggt 2527
QY 2632 TTCTTCAAAAGCTCTAGTATGATGCGCTGTTTATGATTTGGAAGGCTGCTTACCTCTCG 2691
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2528 tttaagaacacaggaacagaagctcggaatttgaagtagccgtctgtgaatccttgcc 2587
QY 2692 ATTCTGTGGTGGGAATTCGTGCAGGGGATTCGAGATTCCTACACTATGATCTC 2751
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2588 ttactatcgggatccgatttgataagaatcagactgccaagaatgcaacgtlacaatcta 2647
QY 2752 TCAGATTTCTTTGTTTCGATGCTATGTAACATCCCAATCTACAGAGCTCTGTG 2811
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2648 actcttggttactatgctggtatctgttcgtatgataccccgactgtaagacaacactgcga 2707
QY 2812 ATGAGCCACAGCTCTGTAAGAAATTCGCGTGGAATCTTTCAGACAGAGCTTTTACTG 2871
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2708 attagcggtatctcttggaacacctcgtagaatttggcaataagacaagctttagtcctt 2767
QY 2872 AGGGGTAGCAACACTACGTGTACACTCCATTTGTAGCTCTTCGACATTTAGCGTATG 2931
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2768 cgtcaggaagacacttcttgtaactcaaatltgaaagctttagccaatttctt 2827
QY 2932 GAATCCGATGATCTTCAAGAACTACAAATGATTTGGTACCAAACTCCGATTTCTA 2990
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2828 gaattggtggtgctatctgcgaattacaatgtagacttaggagacaataaccatctcta 2886

RESULT 13
AAC81914
ID AAC81914 standard; DNA: 273254 BP.
XX
AC AAC81914;
XX
DT 27-FEB-2001 (first entry)
XX
DE Chlamydia pneumoniae genome DNA.
XX
KM Genome; diagnosis; vaccine; ds.
XX
OS Chlamydia pneumoniae.
XX
PN W020002794-A2.
XX
PD 18-MAY-2000.
XX
PF 12-NOV-1999; 99MO-US26923.
XX
PR 12-NOV-1998; 98US-0108279.
XX
PR 08-APR-1999; 99US-0128606.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Stephens R, Mitchell W, Kalman S, Davis R;
XX
XX MPI; 2000-376516/32.
XX
PT Isolated nucleic acid for use in diagnostic and analytical methods
XX
XX encodes genomic sequence of Chlamydia pneumoniae -
```

```
XX
PS Claim 2; Page 128-320; 320pp; English.
XX
CC This invention describes a novel nucleic acid (N1) encoding a Chlamydia
CC pneumoniae protein (P1), given in the specification. The isolated nucleic
CC acid is useful for diagnostic and analytical methods, such as,
CC hybridization-based assays or amplification-based assays. The protein may
CC be used for diagnostic purposes, for their enzymatic or structural
CC activity, or as a vaccine. The invention also describes (1) a probe
CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
CC (N2) that hybridizes under stringent conditions to N1; (3) an expression
CC cassette comprising N1 under the transcriptional regulation of a host,
CC transcriptional initiation region functional in an expression host, and a
CC cassette of (3) as part of an extrachromosomal element or integrated into
CC the genome of a host cell as a result of induction of the expression
CC cassette into the host cell, and the cellular progeny of the host cell;
CC (3) a method for producing a P1 comprising growing a cell of (4) where
CC the protein is expressed and isolating the protein free of other
CC proteins; (6) a purified polypeptide composition comprising at least 50
CC weight % of P1; and (7) a monoclonal antibody binding specifically to the
CC peptide of (6).
XX
SQ Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

Query Match 14.3%; Score 458.6; DB 21; Length 273254;
Best Local Similarity 51.2%; Pred. No. 9.2e-105;
Matches 1534; Conservative 0; Mismatches 1364; Indels 97; Gaps 16;

QY 44 AACCCACTTATATTAATTAATTCCTACTTGGCTCATTAATAAGAAACACAGAGCTCA 103
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98035 aatgcttltgttlltagagaactatcgcatactatagaacaataaagtaaatcaaa 98094

QY 104 AGATAAATTTCTTGAACAGCTGTTTGTCACTTTAATCTTGATTTTCTTTCTTCT 163
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98095 gtaaaagatgacaacaagaagctgaagaatttattctatctcgtgagtttctatt 98154

QY 164 ATATTGATGGGAATGATGCTCTAATAAACAAGACATTACCATGACAGACTTCGATTCCT 223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98155 tataagcgaagaagaatlaataataaagtggtc--targaactcgcaatttccct 98212

QY 224 GGGTTTATGTTTCTCCGCTGTAGCTTCT-----CATGTACACTACAGTCACTACCTA 277
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98213 ggttagtgcctctctgcaacttgcacgttttactagttgttccacgcttlttgcbcaa 98272

QY 278 ACAGAGAACTTTTATACCTGATGATAGCTTTTAATGGAATATGATTCAGAACGTTTA 337
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98273 ctgctgaaataatagagccctctgtagctttagcaggaagtactaacacaggaactata 98332

QY 338 CTCCAAAACTTACAGCCACAAACA-----TATTCCTGTAACAGAGATGCTTCTTTTACG 391
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98333 ctcttaaaataacagcaactcaggaatagatactatctcgaagagagatataactctgcaaa 98392

QY 392 AGCCTGGAAGAGCACTCCCTATCTGACAGTGTGTTTAAAGCAACACGAGACAATCTTA 451
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98393 accttggggatctcgcaagctttaaaggaaggtgttcttctgcacactaaggaactcttaa 98452

QY 452 CTTTCTTGGGGAACGGTCATAGCTTAAAGCTTTATAGATGCTGGACATGATGAC 511
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98453 gcttgcggtlaaggggtactcacttctttaaataaagctc--agtgctgaag 98509

QY 512 GTGCTGCTGATCTACAGCAAGCAATTAAGATCTTACCTCTCAGGGTCTTCTTACTGA 571
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98510 ggcagcaacttctgttaacactgataaataatctgctgcaacaggaatttctgaagctta 98569

QY 572 GTTT-----TGATTCCTCTCTAGACAAACGGTTACTACAGTCAAGCAACGCTTCTCT 625
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98570 ctctcttagcggccctcactcgttaatacaaacccctcaggaagaagtgagcttaaat 98629

QY 626 CAGCAGAGAGCGTAAATTTTGAATAATTTTGTAACTTTGATGTTGCTGGCAATTTTCTA 685
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98630 gtgaggggagtcttaactgtataacaatgaaactatttatttaaacaagattactgtg 98689
```



CC infection. The present sequence may also be used in the  
CC construction of attenuated Chlamydia strains that can over-express the  
CC gene or express it in a non-toxic form.

XX  
SQ Sequence 3000 BP; 871 A; 617 C; 634 G; 878 T; 0 other;

Query Match 14.6%; Score 465.8; DB 21; Length 3000;  
Best Local Similarity 52.0%; Pred. No. 2,2e-107;  
Matches 1372; Conservative 0; Mismatches 1177; Indels 90; Gaps 11;

```
QY 394 CCTGGAAGGACGCCCTTATCTGACAGTGTTTTAAACCAACGACATCTTACC 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 296 ccggaacagcagcagcaatcaaaaagctgtttaacaacactaaggcgattgact 355
QY 454 TTCTTGGGAACGGGTATAGCTTTACGTTGGCTTATATGATGCTGGACTATGAGT 513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 356 ttcaacagtaacggagactctctatgtccaaaagctgagcaggttagcaagg 415
QY 514 GGTGTCATCTACAA--CAGCAATAAGAAATCTTACCTTCAGAGGTTTCTTACTG 570
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 416 gctgcgttaaacagcagcgltgtagtaatactcaacglttaaggtttctctgcta 475
QY 571 AGTTTGAATCTCTCTCTGACACACAGGTTACTACAGTCAAGGAAAGCTT--TCCTCA 627
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 476 tctttatgcgtctcctcgtgaagtcgataactccgcgaaggagcgcttagctgctc 535
QY 628 GGAGGAGGGGTAAATTACAAATATTCGTAACCTGTGTTGCTGGGAAATTTTCTACT 687
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 536 acggtgaagcttgaagttgcaaaaaaatgcaagttgctcttcaagcaaaaacttcaacg 595
QY 688 GCAGATGCTGAGCTATCAAGAGAGCGTCTTCTTAACTGGCACTTCTGGAGATGCT 747
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 596 gataatgycggtgctataccgcgaanaactcttcaatacaggagcaatgltcagct 655
QY 748 CTTTATAGTAACTCTTATCATCAACAAAGGAGACATTTGCTACTACAGCAGCGCT 807
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 656 ctgttttcgaaataatcctccctcaagaagaagcgagccatcgaagcttcgagtcctt 715
QY 808 CGATATGCAATATACACAGGTTATGTTATTCCTATCAATACAGCTACAGTACAGCA 867
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 716 accatlaacggaacccaaggagagctctcttcttctgacaactcttcgcgattctgga 775
QY 868 GGGGCTATGATGATGAAGACGCTCGATATCTATCGAACACAAATTTCTATTTTGA 927
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 776 gctgcgaatttttaacgaagccctcgtgactatttcttaataagctaaagtttccattat 835
QY 928 GGGATG-----CAGCGAAAATCTAGCGCGTGGATC 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 836 gacaataaggtcaacagagcgagctcctcaacaacgaggatatagtcagaggtgctac 895
QY 961 TGCACACACAGGCGAGTGAATCTCTGACATGATATATCAACAATTAAGACTCGATC 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 896 tctgtctataaactcagtaacataactaaggtcaacccctcaggaatacagatglttacc 955
QY 1021 TTTGCTTCAACAGTACAGAAACACAGGCTGGCGCCATCATCTTAAAAAGTAGCCCTT 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 956 ttcaggaacaaatatacagcaacagcgaggtcactatctatgtaaaaaagctcgaact 1015
QY 1081 TCCTCTGGAGGCTTTACAGAGTTTCTACGAATTAATGTCTCATCAG--CAACTCCTAG 1137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1016 gctctcggagagacttaacctatctagtaagaatgltcaatggaagtacagctcctaa 1075
QY 1138 GGGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1197
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1076 ggttgagccatagctatcgaagaatggtggaatgagtttaccgcgcatgltgtgac 1135
QY 1198 ATTATCTTTGTAAGAAATACCTTACACACCGGAATAGCGATATCTTAAACGTAAT 1257
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1136 attgtcttlltagaatacagtaacttctac-----tactccgtgagcagataagaagt 1189
QY 1258 GCGATCAACATGAGAGTAAACGGAAATTCACGGAATTCAGGGCTGCTTAAAAATCTATA 1317
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
DB 1190 agtataccttaggaagaggtgcaaaagatgacaagcttgcgttcgtcgtgtagagcc 1249
QY 1318 ATTTTCTTATGATGCCATCTTCACTTACAGAGAACCTCA-----TCAGACTATTTGAG 1371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1250 atctactctatgatacccaactaactaactaactaactaactaactaactaactaacta 1309
QY 1372 ATTAATTAACGGCTTCGGGAGCTCTCAATTCATATCAAGAAAGCAATTTTCTGGA 1431
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1310 gttatgagatcccgagattcttcgactacaataataaggaacatctcttcaacagga 1369
QY 1432 GAACCCCTTACAGCAGATGCACTTAAAGTTGCTGACATATTTAAATCTTCAATTCAG 1491
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1370 gaaagattatcagagacagagcgccagatctctaaatacttactactcgaagctacac 1429
QY 1492 CCAGTCTCCCTATCCGGAGAGAAAGTATGTGTACAAAAGGAGTCACTTAGAGGACAG 1551
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1430 ccgttaaccccttcagaggtactctactcttaaaacagagaggtcctcgaactcag 1489
QY 1552 AGCTTCTCAAGAGAGCCGTTCTCTCTCGGAGATGATTCAGGAACGACATTAATCACT 1611
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1490 gatttaactcaacagcgagattctcgtctcgaaatgagcgttagaactactctagagact 1549
QY 1612 ACAGCTGGGAGATTTATCAATTCACGAACCTAGGAATCAATGTGACTCCTTAGGTCTTA 1671
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1550 --gctgataactagacccaataaattgtcatataacataagctatagaaggtgca 1606
QY 1672 CAGCCGCTCAGCTTAAACAGCAAAAGGCTGCTCAAAATTAAGTATGATCTGGAAGCTC 1731
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1607 aagaagggcaaaaatagaacccaagctagcgtcaaaaaactgaacttacttgaacacatc 1666
QY 1732 AACGTGATTTGATTTGAAAGGAAACATTTATGAAGTCAATATGTTCAACCATCAAGCTC 1791
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1667 acttattgagcccgagcgagcgatctttagaaatacatagtttaagaatcctcagctc 1726
QY 1792 TTCTCTCTTTTAAAAATCAACGCTGATGCTGATGTTGATATCAATACAGTTTACATCAG 1851
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1727 taagacatcttagagctcaaaagctcttgaaactgtaacaagacgcag----- 1775
QY 1852 CTATTCCTCTGTCCTGCTGAGATGCTTAAATTCAGAAATTAAGTATGATCTGGAAGTAT 1911
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1776 -tgactcagatccctataataggtgtagaaatccatcctacaggtcaltcaggaactg 1834
QY 1912 GTTAAATTGACATACAGATACAGCTACCAATACAAAAGAGCCACGGCACTTGGACCA 1971
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1835 ccaattgtttggggagcagggctcttctacgac-----tgaaccttcaactcggactaa 1888
QY 1972 ACAGATTTGTTTCCACGCCCGCAAGAAAATCTGCTTATGTTATGCAATATCCTATGGGGA 2031
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1889 actggtatattccttaactccgagcgtaacgctctttagtcccttaactagctatggaat 1948
QY 2032 GTCTTACTGACATTCGCTCTGCAACAGCTTGTAGAGATGCGGCAACTGTATGAA 2091
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1949 gcaattatagataatagctctctccatctatctataggaactcgaacgaagaggtgacg 2008
QY 2092 CACAAACAAAGTTTCTGAGTTTCTCCATAGCAACCTTCTCATTAAGCTGAGATGAA 2151
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2009 ggaagacggtcttttgggtgctgattatcactccttccataagaatagtaaca 2068
QY 2152 AATCCGAAGGCTTCCGATACCTCTGAGAGCTACGATGAGTGGGAAGTCTCACACT 2211
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2069 aacagcagcggtcttcgcatcttagtgaggtggtatgtaagtaagaacactaactact 2128
QY 2212 CCTAAAGACAGCCTATTTATCTTTGGGTTCTCCATCTCTCTTGTAGACACAAAGTTGT 2271
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2129 tgttcaagaataatcttctgctgcatcttcttgaggtcttggagagatagagactac 2188
QY 2272 TTTATGCTCAACAACTCTTGAACCTTACGTTGGAACCTTTATCTTCAAGCACTCTCAT 2331
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2189 ttgttagctaaagaatcaaggtacgtctcagaggaactctatatacagaacaacgaa 2248
QY 2332 ACCGTACAAACCCCAAAACATTTGAGATTAGGAAGGCAAACTTTTCTGAATCAGGTATA 2391
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2249 acctatattctctcctcgtgcaactaagcgctgtgtcgttctatag----- 2297
```

```

Db 1507 aagaagcaaaaatagaaaccaaagctacgtcaaaaaatctgacttatctggaaccatc 1566
QY 1732 AACCTGATTATATGGAAGGACATTTATGAAGTCAATGTTCCAGCCATGACCACTC 1791
Db 1567 acctatggaaccgagcgacgtttatgaaatcatatgattaaagaatccctcagttcc 1626
QY 1792 TTCTCTATTAATAAATACAGGTGTGATGCTGATGTTGATCTAAGCTTGACATCAGACG 1851
Db 1627 taagacatcttagaagctcaagaagctctggaactgtaacaaacacgcagag----- 1675
QY 1852 CTATCCCTGCTGCTGCTGAGATCTCAATTCAGAAATACGATTCAGATTCAGCAATGGAAT 1911
Db 1676 -tgaccagatcccttaataatggtgagaatctccattacggtcatcagaagaaacttggg 1734
QY 1912 GTTAATGAGCTACGATACAGATACAAATACAAAGAGCCAGCCAGCACTTGAGCAAA 1971
Db 1735 ccaattgttgggagagaggtctctacgac-----tgcaaccttcaacttgcactaaa 1788
QY 1972 ACAGGATTTGCTCCAGCCCGGAAAGAAATCTGCCATTAGATGACATACCTATGCGGA 2031
Db 1789 actggtatattcccttaaccgagcgatcgcgtctttagtcccttaataagctaatggaat 1848
QY 2032 GTCTTTACTACATTCGCTCTCTGCAACAGCTTGATAGATCGCGCACTGCTGATGGA 2091
Db 1849 gcaattatagatatagctctctccatctatctatgagagctgcaaacgaggtgctcag 1908
QY 2092 CACAAACAGGTTTCTGGGTTCTCTCCATGACGACACTTCTGCAATGAGACTGGAGATGAA 2151
Db 1909 ggaagacgtcttcttgggtgtgtgtggtacgtacactcttcacataagatagatcaaaa 1968
QY 2152 AACGCAAAAGCTTCGCTACACTCTGAGAGGCTGAGCTGAGCTGAGCTGAGCTGAGCT 2211
Db 1969 aacagagcggtgtctgccccttggagtggtggtatgacatagagagaaacctacataact 2028
QY 2212 CCTAAAGCAGCCTATTATTCCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2271
Db 2029 tgttcagataagattctctagtgctgcatcttctgacgtcttctggaagagatagagactac 2088
QY 2272 TTTATGCTCTACACAACTATGACACTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2331
Db 2089 ttctgtagctaaagatacgaagtagctacgaggaactctcattacacagcaagaa 2148
QY 2332 ACCCTACAAACCCCAAAACCTATTGAGATAGGAAAGCAAGTTTCTGATCAGTATATA 2391
Db 2149 acctatatactctctctctctgcaactacgagcctgttctgctctatg----- 2197
QY 2392 GAAAAATTCCTCGGAGGAAATTCCTCCCTAGCCTTGAGATGCCAAGTTTCTGAGCCATCA 2451
Db 2198 -----ttcctacagagattcctctgtctcttctcaggaacacttagctacaccatacg 2250
QY 2452 GACAAACGCTATGAAAGGCACTATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2511
Db 2251 gataacgagatctgaaacccaacatatacactactcgtttaaaggaagctggggaat 2310
QY 2512 GAGCTATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2571
Db 2311 gataagttcgtctttagaattcgttggaagagctccgatttgccttagatga---aagtgtc 2367
QY 2572 CTATTCAAGACCTTCATTCACAGATGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2631
Db 2368 ctatttgagagagatcagtcctcctcatgaatttgcatttgcataacatcaggaaggt 2427
QY 2632 TTCTTTGAAAGCTCTAGTATGAGGCTGTTTATGATGGAAGGCTGTTTAACTCTGCG 2691
Db 2428 tttaaagaaacaggaacagaagctcgtgaatttgaagtagcgcgtcttgcgaatcttgc 2487
QY 2692 ATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2751
Db 2488 ttaccatctcgagatccgatttgaataagaaatcagacgccaagatgcaacatcaatca 2547
QY 2752 TCAGAGATTTCTTTCGATGCTATCTATACAAATCCCAATCTACAGGCACTCTTGTG 2811
Db 2548 acctctggtataactcgttgagctctgttgcgtagtaaaccccgacgtgaacaaacatg 2607

```

```

QY 2812 ATGAGCCCACTCTTGAAATTCGCGGTGGCAATCTTCAAGACAGCACTTTTACTG 2871
Db 2608 attagcggtagctctcgtgaaacaccttcgtagcaagattgcaagaagctttagctct 2667
QY 2872 AGGGGTGACCAACTCTGCTCAACTCCCAATTTGAGCTCTTGGCAATTCAGCTATG 2931
Db 2668 cgtgcagggagacatttcttgaactcaaatlttgaagcctttagccaatlttctt 2727
QY 2932 GAATCCGCTGATCTTCAAGCAACTACATGATGATGTTGTTACCAAACTCCGATTCTA 2990
Db 2728 gaattgcgtgcatctcgcataactaactagactagagagcaaaatcaacttcta 2786

RESULT 12
AA27021
ID AAA27021 standard; DNA: 3000 BP.
XX
AC AAA27021;
XX
DT 11-AUG-2000 (first entry)
XX
DE Chlamydia pneumoniae 98kD putative outer membrane protein gene.
XX
KW Chlamydia; antigen; vaccine; infection; outer membrane protein; ds.
XX
OS Chlamydia pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 101..2887
FT /tag-a
FT /product- "98kDa putative outer membrane protein"
XX
PN W0200026237-A2.
XX
PD 11-MAY-2000.
XX
PF 29-OCT-1999; 99MO-GB03579.
XX
PR 29-OCT-1998; 98US-010670.
PR 01-MAR-1999; 99US-0122066.
PR 27-OCT-1999; 99US-0428122.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Mordin AD, Oomen RP, Dunn PL;
XX
DR WPI: 2000-365569/31.
XX
DR P-PSDB: AAY94327.
XX
PT Novel Chlamydia 98 kDa putative outer membrane protein antigen, used
PT for vaccination and protection against Chlamydia infection
XX
PS
XX
Claim 1: Fig 1: 93pp; English.

The present sequence is the 98kDa putative outer membrane protein gene
from Chlamydia pneumoniae. The genomic sequence was amplified using two
PCR primers. The 5' primer contains a NotI restriction site, a ribosome
binding site, an initiation codon and a sequence close to the 5' end of
the 98kDa putative outer membrane protein coding sequence. The 3' primer
contains the sequence encoding the C-terminal sequence of the putative
outer membrane protein and a BamHI restriction site. The stop codon was
excluded and an additional nucleotide was inserted to obtain an in-frame
C-terminal fusion with the Histidine tag. The PCR product was cloned
into a eukaryotic expression vector (pCA-Myc-His) by restricting both
the vector and the PCR product with NotI and BamHI and performing a
ligation reaction. This expression vector was injected intramuscularly
and intranasally into mice, which were subsequently inoculated with
Chlamydia pneumoniae. The chlamydial lung titers of the immunised mice
were lower than those of the controls. Thus the 98kDa putative outer
membrane protein can be used as a vaccine to provide protection against
Chlamydia infections, especially Chlamydia pneumoniae infections.
The polypeptide may also be administered orally to treat Chlamydia

```



QY	1068	AAAGCTACCCCTTCCTCTGTGAGGCTTTACAGAGTTCTTACGAAATATGTCATC---	1124
Db	885	aaaattggtctatccctccaggaagagaggtgtattcttctaacaacaaagctgcgaa	944
QY	1125	AGCAACCTCTTAAGGGGGGTGCTATACACATCGATGGCTCAGGAGAGTCAGTCTTTCGC	1184
Db	945	tgcatactctaaagagagggaattgsgatttcgaattcttgagagatlaagatcttcgc	1004
QY	1185	AGAGACGAGAAACATTAACCTTTGTGAAGAAATACCTTACACACACCGAAGTACCGATAC	1244
Db	1005	agatctcggaatcatcatcttcgagggaactacgaagcactaaggaagtcctgcgag	1064
QY	1245	TCTTAAGCTTAATCGATTCACAAATAGGAAGTAAAGGGAATTAACGGGAATTACGGGGTGC	1304
Db	1065	tgtgcacgaagaaagtcatagatcttgatcgaatgcgaatgcaaaattttaaactctccgagac	1124
QY	1305	TAAAAATCATATCAATTTTCTTCTATGATCCACATCTTACAGAGACACTCATAGACGT	1364
Db	1125	tcggggaaataaagtatttctcatgatccatcagagctcaga---gctactgataa	1181
QY	1365	ATTGAAGTAAATTAACGGCTCTCGGGAGCTCCAATTCATATCAAGACGATTTATTT	1424
Db	1182	gctctcttgataaagctgcagcagagatctgaaatacctatggaagctaatctgatt	1241
QY	1485	TTTCGAGAGAAACCTTAACACAGATCAACTTAAGTTAGTGTGACAAATTTAAATCTTCAT	1484
Db	1242	ctctgagagaaactctccagagaaggaacttaagaaacctgcaactcgaagctcaact	1301
QY	1485	CACGACGCACTCTCCCTATCCGAGAGAAAGTTATTTGCTACAAAGGAGTCATTTGA	1544
Db	1302	taccaaagctgtlaagatctgcacaggtgcctagattgaaagaatggaagtacgtagt	1361
QY	1545	GAGCAGCAAGCTTCTCTAATAAGAGCCGGTTCTCCCGGCATAGATTAAGACACACTT	1604
Db	1362	tgcataactcataacgaaggtcgaaggatccgaagtcgltatggaatggaaggaactactt	1421
QY	1605	ATCAACTACAGCTGGGAGATTTCATATCACAGAACTTAGAATTCATGTTGACTCCTTAGG	1664
Db	1422	tgaagcaagcgtcgaagggggtcactctcaatctgagcctacgaatataagatctccitga	1481
QY	1665	TCTTAAGAGCCCTTCAGGCTTAACGAAAGGTGCTTCAATTAAGATGATGCTATCGG	1724
Db	1482	tgggaacaaataaagctctcatcctaagcgaaggaagaagtaagagatgtgcctatacgg	1511
QY	1725	GAAAGCTCAACCTGATTCGATTTGAAGGAGAACATTTATGAAGTCACTATGTTCAAGCATGA	1784
Db	1542	gctcatcatgctgtlagaatctcgaaggaactatratgsgcataactccagtcaca	1601
QY	1785	CCAGCTTCTCTCTATTAATAATTCACAGGTGATGCTGATGTTGATTAACGTTGACAT	1844
Db	1602	gcagatctcttccttaataagatcttctgcacaaggaacgaatgaactaactaagatctcc	1661
QY	1845	CAGACAGCTTATCCCTGTTCTCTGCTAGAGATCCTTAATTCAGAAATACGATTCCAAAGACA	1904
Db	1662	cgataccccaat-----tctaataactaagatcatcctatggtatccaagaac	1709
QY	1905	ATGGAATGTTAATTTGACATACGAGTACAGCTACAAATACAAAGAGCCACGCAACTTG	1964
Db	1710	tgaataatgttgttgggtcgaagatgcaactgcacaacaaatgttactcttaactgt	1769
QY	1965	GACCAAAACAGGATTTGTTCCAGCCGCCAAGAAATAATCTGGGTTAGTATCAATTAACCT	2024
Db	1770	gactcaaaacggtatacaagccgaatccagaagctcaggggaacttggttcctaataagct	1829
QY	2025	ATGGGAGTCTTTACTGCAATTTGGCTCTCTGCAACAGCTTGTAGAGATCGGCGCAACTCG	2084
Db	1830	gtggggtctcttgttcgaatgctccgcacatccgaagcctcatgscgagcacaaagtctc	1889
QY	2085	TATGGAACACAAACAGGTTTGGGTTTCTTCATGACGAACCTTCTGCATTAAGACTCG	2144
Db	1890	gtatactctgtcaacaatattgtgggtatcaagaatccggaactcttttgatgaatgaatca	1949

[illegible]

DE Chlamydia pneumoniae surface exposed protein Omp8 DNA.  
XX  
XX  
Omp8, outer membrane protein 8; surface exposed protein; antigen.  
KM Infection; diagnosis; vaccine; atherosclerosis; asthma; ss.  
KM



OY	477	AACGTTTGGCTTTATAGATGCTGGCACTCATGACAGTGCCTG---CATCTACACAGC	533
Db	394	tcattcttgcacaattatcttctgctactgtagcaggtgtgtgtgtatgaatacacaagc	453
OY	534	AAATTAAGAAATTTCACCTTCTCAGGCGTTTCCCTACAGAGTTTGGATTCCCTCCAGAC	593
Db	454	ttctgggaattccgaattctccagaattttcaactcttcggatgcttcgagctcccaag---	510
OY	594	AACGGTTACTACAGTCAGGAAACGCTTTCCTCAGCAGGAGGCGTAAATTTAGAAAATAT	653
Db	511	-----gaccacagytlaagagagccattaaattaccagatgctcgtgtgtttgagat	564
OY	654	TCTGTAACCTTTACTTGCCTGGGAATTTTCTACTGCAGATGGTGGAGCTATCAAGAGAC	713
Db	565	aggsaattcttgatcttaattgaataatgctctatgtgaataatgggggagccataatacga	624
OY	714	GCTTTTCCTTTTAACTGAGCCTTCTGAGATGCTCTTTTATGAACAACTCTTCATCAAC	773
Db	625	gactttgctcttcgcctcggagaglaagcggttttgacgcttcccttggaataagctcgcga	684
OY	774	AAAGGAGAGCAATTTGCTCTACTACAGCAGCGCTCCGATAGCAAAATPACAGGTTATGT	833
Db	685	acaaaggggagcgatctatgctctctcgtgctcgtgattctcgagaatgcaggaactct	744
OY	834	TAGATTCTATCTAACATPACGCTCTACGTACGTACGAGAGCGCTATCATGATGAAGCACTG	893
Db	745	gagctctcggaaacaaacagtcgcgaacacatcaagagcgagctctcgtcgaaggaaact	804
OY	894	GATCTATTCGAACAACAAATTTCTATTTTGAAGGAATGCGCCGAACAACTPACGGGG	953
Db	805	tgtatctcccaatacacaataatcttcttcgltcgtgcgtcgaagcaactacaatcggcg	864
OY	954	TGCGAT-----CTGCAACACCAAGCGAGTGGATCTCTGAACTGATTAATCTTAACAA	1007
Db	865	agctatgtattgtaacaaagcagggggaacccagaccctatcttgactcttcaaggaa	924
OY	1008	TAAAGCTCTGATCTTTCGCTTCAACAGTAGACGAACAAAGCGTGGCGCATCTCATGCTTA	1067
Db	925	tgaagacctgcatcttcttgatacaacaagaagaaatagtgaggtgcatttataccea	984
OY	1068	AAAGCTAGCCCTTTCCTCTGAGAGCTTTCAGAGTTTTCAGCAAAATTAATGCTCATC---	1124
Db	985	aaaattggtgtatccctccaggaacggagggaggtgttatttcttaacaacaagctcgcga	1044
OY	1125	AGCAACTCTTAAGGGGGGTCTATCAGCATCATGCGCTCAGAGAGCTCATGCTTTTCG	1184
Db	1045	tgtactccctaaggaggggcaattcgatctagttcttgagagattagatattctgc	1104
OY	1185	AGAGACAGGAACCTTACCTTTGTAAGAATACCTTTCACACACCGCAAGTACCGATAC	1244
Db	1105	agatctcgcgaatcatcttctcgaggcaatactacgacactacaggaagtcctcgag	1164
OY	1245	TCTTAACGTAATGCGTTCACAACTAGGAAGTACGGGAATTCACGGAATTCAGGGCTGC	1304
Db	1165	tgtagcagaatgctatagatcttgcatacgaaatgtgaatllttaaatcttcgcaggac	1224
OY	1305	TAAAAATCATATCAATTTTCTTATGATCCCATCTTTCAGAGAACCTCTCATCAGAGT	1364
Db	1225	tcgggggaataaagtattcttctatgatactcatcaagagctcaaga---gctactgataa	1281
OY	1365	ATTGAAGATTAATTAACGCGCTTCGCGGAGCTCTCAATTCATTCAGAGAACGATTTCTAT	1424
Db	1282	gctctcttgataaagctcgaagcagagatcttgaaatacctaataagagctacatacgctt	1341
OY	1425	TTTCTGGGAACACCTTAACAGCAATGAACTTAAAGTGTGTGCAATTTAAATCTTCTAT	1484
Db	1342	ctctcggagaaactctcagaagagaaacttaagaaacctgcgaactcgaagctctacat	1401
OY	1485	CACGACACCGTCTCCCTATCCGAGGAAGAAATTATGTCACAAAAGGGGACACTTTTGA	1544
Db	1402	tacacaggtcgttagaggtctgcgcaggtgcctctaglattgaaagatgagtgagatcgtagt	1461
OY	1545	GAGACACAGCTTCTCTCAAGAGCGCGTTCTCTCCTCCGCAATGAGTTTACAGACACAT	1604

Db	1462	tgcaatactctaaocgaagctcgagggaatcgaaagtctgttaaggatggaggagactctt	1521
Qy	1605	ATCAACTACACCTGGAGATTTACAAATCCAGCAACTAGCAATCAATGTTGACTCCATTAG	1664
Db	1522	tgaaggcaagcgctgaagggtctcatctccatctgacctagccataataatagatcttccctaga	1581
Qy	1665	TCTTAAGCAGCCCGCTCAGCCTACAGCAAAAGTGCTTCAAAATAAAGTATGTATCTGG	1724
Db	1582	tgggacaataaagctatcatctaaaggagcgagcgaagaagatgcttgcccttcagg	1644
Qy	1725	GAAGCTCAACCTGATMTGATATTTGAAAGGAACATTTATGAAAGTCATATMTCCAGCATCA	1784
Db	1642	gactatactgtctgaatgtctcaggggaaactatgaagcaatcaatctccagtcgaaca	1701
Qy	1785	CCAGCTCTCTCTGTATTTAAATAACAGGTTGATGCGTATGATCTATACGTTGACAT	1844
Db	1702	gcaggtcttctcttaataagacctctctgcacaagaaagatgacctactaagatatccc	1761
Qy	1845	CAGCAGCCCTATCCCTGTCTCCCTGCTGAGATCTTAATTCGAATPACGATTCGAAGACA	1904
Db	1762	cgataccccaat-----tctaatactacgaatcactatggtatcgaaggga	1801
Qy	1905	ATGGAATCTTAATTGGACTACGATACAGCTACAAATACAAAGAGCCGCAACTTG	1964
Db	1810	ctggaatactgtcttggtgcagcagatgcaactgcaaaaaaataatgctactcttaacttg	1865
Qy	1965	GACCAAAACAGAGATTTGTTCCACACCCCGCAAAATAATGCGTTGATATGCAATACCT	2024
Db	1870	gactaaaaacaggaatacaagccgaatccagaagctcaggaccttggcttccataagct	1922
Qy	2025	ATGGGAGCTCTTACTGACATTCGCTCTGCAACACTTGTAGAGATCGCGCAACTGG	2084
Db	1930	gtggggtctcttcttgcatgtcgcctccatctcagcctcaatgacccgaagcagaacttc	1988
Qy	2085	TATGGAACACAAACAGGTTTTCGGTTTCTCCATGACGAACTTCTCGATTAACAGCTGG	2144
Db	1990	gtctatctgtcaacaatatttggtgataagaaatcggagacttctgcatgaagatca	2048
Qy	2145	AGATGAATAATGCAAAAGCTTCCGCTACCTCTGGAGCCTGATCGTATGGTGAAGTGC	2204
Db	2050	gaaaggaaaccaaagtgtatgtcatctttagcgggggtatgcatctagggggagatt	2108
Qy	2205	TCACACTCTTAAGAGCAGCCTATTATTCCTTGGCTTCGCCATCTCTTGGCTAGAGACA	2264
Db	2110	cttcacggctcttgaaaattcttctaatttgcatttgcagcttlttgctacgacaa	2165
Qy	2265	AGATTGTTTATCGCTACCAACAACACTAGAACCTACGGTGAACCTTATTTCTTCAACA	2324
Db	2170	ggaccatcttctgcttaagaacacatccatctatctatgcaagggcaatgaattaccgaca	2228
Qy	2325	CTGTCATTCCTTACAAACCCCAAAACATTTTGAGATTAGGAAGCAAGATTTTTCGATTC	2384
Db	2230	c-----ctcggagagatctaagaacctctgctaa	2255
Qy	2385	AGCTATGAAAAAATTCCTTAGGGAATTCCTCCCTACGCTTGGATGTCGAAGTTTCGTTGAG	2444
Db	2257	gatttctaggaataatcttgacctccactacttctgtcttcaatgctcgglttgctatag	2318
Qy	2445	CCATTTCAGAACCGTATGGAAGACGACATATCTCATTTGCCAGAAATCCGAAGTTCTTG	2504
Db	2317	ccatacgaacataacatgaccccaagaatcacatctgctatctccctgttaaaggaaagct	2378
Qy	2505	GAGCAAGAGGTATPACTGTGTGGTATCGGCCATAGACCTTCCTTTTGTCTTTTCCAAACC	2564
Db	2377	gggaaatgaatgcttgcgtatagaatgtggaggagctatcccgtagtcttcaggagc	2438
Qy	2565	ACATCTCTTTTTCAGAGCTTCATTCACAGATGGAAGTCGAATAGTTTATGTATTCACA	2624
Db	2437	tcggtcttggttgatctcccaacgcgcatttctaaccctcaggatgactatgacatca	2498
Qy	2625	AAATAGCTCTTCTGAAAGCTCTGTAGTGGCCGTGGTTTACTATTTGGAAGCTGCTTAA	2684



```

Db 1808 taatccgagcagagagcttcgtagtcccgatcattatgagcattactgacat 1867
QY 2046 TCCCTCTCTGCACACAGCTGTAGAGATCGCGCACTGGTATGGAACACAAACAGTTT 2105
Db 1868 tcgactctacagcagatcatgacatcgcaggaataglatctacagcaagagagact 1927
QY 2106 CTGGGTTTCCCTCCATCAGCAACATTTGCTGCAAGACTGAGATGATAAATGCAAGGCTT 2165
Db 1928 ctggagcatcagagacgcgaattctctccataagataatcaggaactaaccaagcatt 1987
QY 2166 CCGTCACTACCTCTGGAGGCTACGTCATCGGTGGAAGTCTCAGACTCTCTAAGAGCAGCT 2225
Db 1988 ccgacataaagaagtaagcatatgtgtgaggaagtgctgaagaattctctgaaatat 2047
QY 2226 ATTACCTTTGCTGCTGGCCATCTTTGCTGATAGACAAAGATGTTTATGCTGCACAA 2285
Db 2048 ctccagctgtagctctctccagcctctcgtaagaataagagcgtctctatagtaaaa 2107
QY 2286 CAACCTAGAACCTACGAGTGGAACTTTATCTTCAAGCACTCATACCTACACCCCA 2345
Db 2108 taactctataactatattagcgtcgtatatacctgcaacatcgagcattcctagagagact 2167
QY 2346 AACTCTTTGAGATTGGAAGACCAAAATTTCTGATACGCTATGAAAATTCCTCTAG 2405
Db 2168 tcccatgacctctcttgagaag-----tataccgacatgctgaa 2206
QY 2406 GGAATTTCCCTGAGCTGTGGAGTGCAGATTGCTGACGCAATTCAGACCGTATGGA 2465
Db 2207 agatctccctcatttctgaagccagccagctacagctacacaaatgatatgaa 2266
QY 2466 AACGCACTATACCTGATTCGAGCAATTCGAAAGTCTTTCGAGCAAGAGTGTATACCTG 2525
Db 2267 tactcgctatactctcctatccctgaagctcaagcctcttgaccataactctgagagctct 2326
QY 2526 TGGTATCGGCTCTGAGACCTCTTTTCTTCTTCCAAACCAACATCTCTCTTCAAGACTT 2585
Db 2327 agagccctcgagagatcctcgtctatatactccctaaagaagccctctctccaagagata 2386
QY 2586 CATTCACAGATGAAGTGAAGTGAATGTTATGATATCATCAAAATAGCTTCTCGAAAGCTC 2645
Db 2387 ttccctctcttaagaagttccagcagctctacagccgcacaaacttaagaagagctg 2446
QY 2646 TACTGATGGCCGTGTTTATGATTTGGAAGGCTGTTAACCTCTGATTCCTGTTGGTGC 2705
Db 2447 cgtctgagagcccgctcttgatgatgagacctaagcctgacgtctctacccgtcgcgat 2506
QY 2706 GAATTCGTGCAAGGGGATATGCGAGATTCTACACCTATGATCTCTCAGAGATTCTTGT 2765
Db 2507 tcggtctagaataaaatctccgaagatgaaaaataatttcgagattctctagcctacat 2566
QY 2766 TTCCGATGTCATCTGTAACATCCCAATCTACAGGAGCTCTGTGATG--ACCCAGAC 2823
Db 2567 tggtagtgatgatacgtataaaatcccgcttcgactctctctatgactcagtgagagcc 2626
QY 2824 TCTTGAAAAATTCGCGGTGGCAATCTTTCAAGACAGGCAATTTTACTGAGGGGTAGCAAC 2883
Db 2627 tcttgacttcgcttagtaaaaaacctcgcaagcagcctctcttagcaagtgctggaagc 2686
QY 2884 AACTAGCTTACAACTCCCAATGTGAGCTCTTCGACATTTAGGCTATGGAACCTCCGTGA 2943
Db 2687 ca-tctgactctctccctatctagaaactctctggggaagcctctatgagctctgagc 2745
QY 2944 TCTTCAAGAGACTACAAATGATGTTGGTACCAAACTCCGATTC 2998
Db 2746 tcagcacacatctacaatgtatgtggtggaagtaagatactcattc 2790

```

```

RESULT 9
AAZ61509
ID AAZ61509 standard; DNA; 2957 BP.
XX
AC
XX AAZ61509;

```

```

DT 19-JUN-2000 (first entry)
XX
DE DNA encoding the CPN100395 polypeptide.
XX
KW CPN100395; Chlamydia infection; immune response; vaccine; ss.
XX
OS Chlamydia pneumoniae.
XX
FH Key location/Qualifiers
FT CDS 101..2857
FT /ftag= a
XX
PD WO200011183-A2.
XX
PF 02-MAR-2000.
XX
PR 18-AUG-1999; 99WO-IB01449.
XX
PR 20-AUG-1998; 98US-0097187.
PR 20-AUG-1998; 98US-0097188.
PR 20-AUG-1998; 98US-0097189.
PR 20-AUG-1998; 98US-0097190.
PR 20-AUG-1998; 98US-0097195.
PR 20-AUG-1998; 98US-0097196.
PR 20-AUG-1998; 98US-0097197.
PR 27-AUG-1998; 98US-0097191.
PR 17-AUG-1999; 99US-0376770.
XX
PA (CONN-) CONNUGHT LAB LTD.
XX
PI Murdin AD, Oomen RP;
XX
DR WPI: 2000-224703/19.
XX
P-PSDB: AAT69369.
XX
PT - Novel antigens and corresponding DNA molecules that can be used to
PR prevent, treat and diagnose disease caused by Chlamydia infection in
PT mammals, especially humans -
XX
PS Claim 1; Fig 15-E; 201pp; English.
XX
XX AAZ61502-09 encode Chlamydia pneumoniae polypeptides. The polypeptides
CC are present in the bacterial membrane structure, in the external
CC vicinity of the membrane structure, in the inclusion membrane
CC structure, in the external vicinity of the inclusion membrane structure,
CC and in the cytoplasm of the infected cell. The polypeptides may be
CC used to prevent, treat and detect the presence of Chlamydia infection
CC and/or the presence of Chlamydia in a sample. The polypeptides may
CC also be used to induce an immune response in a mammal. The vaccine
CC vector comprising the polynucleotides is used to induce an immune
CC response in a mammal. Antibodies directed against the polypeptides
CC may also be used therapeutically to treat and/or prevent a Chlamydia
CC infection.
XX
SQ Sequence 2957 BP; 818 A; 598 C; 683 G; 858 T; 0 other;

```

```

Query Match 15.9%; Score 508; DB 21; Length 2957;
Best Local Similarity 51.8%; Pred. No. 5,1e-118;
Matches 1401; Conservative 0; Mismatches 1230; Indels 75; Gaps 8;

```

```

QY 297 TGATGATGCTTTAATGAATATGATTCAGGAGCTTACTCCAAATCTCACCCAC 356
Db 214 tggtagtaacaagcacaagaattactctctaaagggcaactctcgatgctagtgagc 273
QY 357 AACATATTCCTTAACAGAGATGCTCTTTTACGAGCCTGCAAAAGGCACTCCCTATTC 416
Db 274 gacctatattctcgatgggagctcgtcgaataagcagaaggaagaaacaaagcctaac 333
QY 417 TGACAGTCTTTTAAGCAACACGAGCAATCTTACTCTTGGGGAAGAGGTATGCTT 476
Db 334 caaagtgcttctctacaacacgcaggaatctctactctcttagggaagcagattctct 393

```





D	1338	ctcggnaaccttagcaactcaagaagaaigtgcgaagttagtgcatacgtgttccaacagac	1397
Q	1566	GGCCGGTCTCTCTCTGGCATGGATTTCAGGAACGACATTATCACTACAGCTGGAGAT	1625
D	1398	tgaaggtctctacactctccctcgtacccaagaacaaagctcaaaagcagatactcgaagct	1457
Q	1626	TACAAATACCAACCTTAGGAATCAATGTTACACCTTAGTCTTAAAGCAGCCGTAGCT	1685
D	1458	cagctctacaacactgctgtgtaactcttcctgccttagagggaaataagaatgtgtccat	1517
Q	1686	AACGCAAAAGGTGCTTCAAAATAAAGTATCGTAATCTGGGAAGCTCAACCTGATGATAT	1745
D	1518	tgaacacgacgagccaacaaactaactaactcctcctctgtttcccaagatag	1577
Q	1746	TGAAGGAAACATTTATGAAGAAGTCATATGTTTCAGCATGACCAACGCTCTCTCTATTA	1805
D	1578	taagcggaattctttaaagaagccaatacgcataaagaagccttcacgcagccttggagt	1637
Q	1806	AATCACGGTTGATGCTGATGTTGATCTAACTAACCTTGACATCAGACACCTTATCCGTTC	1865
D	1638	attcac---tgcctcactgcctgcctagagatatattatcagtgctcttcacactcc	1694
Q	1866	TGCTGAGGATCTTAATTCAGATACGGATTCCAGAGACATGGAATGTTAATTGACTAC	1925
D	1695	agtaacaaccccaacccatcaactacgagatcaggaacatctggaaacacacttggcaga	1754
Q	1926	GGATACAGCTCAAAATACAAAGAGGCGCACGACCTTGACCAAAAGATTTGTTCC	1985
D	1755	caatacaactgcgaactca-----ggaactatgacttgggttaactaagcgctacaacc	1808
Q	1986	CAGCCCGGAAAGAAATCTGCGTTAGTATGCAATACCCATGAGGAGTCTTACTGACAT	2045
D	1809	taatcccgagcgtagagctctcgttaqtlcccgatccatattatgycatcccttactgcat	1868
Q	2046	TTCGCTCTGCAACACTTGTAAGATGCGGCACTGCTATGAGAACCAAAAGATTT	2105
D	1869	tcgaacctcaacgaagatcaatgacatcctaagcgaatagatatactacgaacaagagact	1928
Q	2106	CTGGGTTTCCCTCATGACGAACCTTCCCTGCATTAAGACTGGAGATGMAAATCCGAAGGCTT	2165
D	1929	ctgggacatcagagactctgcaattcttcataagataaatacagaactaaccgaagact	1988
Q	2166	CCGTCATACCTCTGGGAGGCTACGTCACTCGGTGGAGTGTCTACACTCCTAAAGGAGCT	2225
D	1989	ccgacaataaagaactgaagctatactgttgagaaggtctgtaagaatttcttcgaaatat	2048
Q	2226	ATTTACCTTGGCTCTGCGCATCTTCTTCTGAGACAAAGATTTGTTTATGCGTCACAA	2285
D	2049	cttcagtgtagcttcttcgcagactcttcggttaagaataaagccgtttatagtgtgaaa	2108
Q	2286	CAACTCTAGAACCTACGCTGGAACCTTTATCTTCAAGCATCTCTACCTACACCA	2345
D	2109	tactctcaataactatttagcgtgcgtatacctgcaaacatcgagacttcccttagagagact	2168
Q	2346	AAACTATTTGAGATTAGGAAGAGCAAAAGTTTCTGATACGCTATAGAAAAATTCCTAG	2405
D	2169	tcccatgcctcctaatttggaaag-----taccacgcaatgctgaa	2207
Q	2406	GGAATTTCCCTAGCCCTTGATGTCCAAATTTCTGTTACAGCCATTTCAGAACCCGATGGA	2465
D	2208	agatattcctctcatatttggatgcccagtgtaagcgtctacactaaataatgatatgga	2267
Q	2466	AACGCACTATACCTCATTTGCGCAGAAATCCGAAAGTTCTTGGAGCAACGAGTATAGCTGG	2535
D	2268	tactcgtatactactctactaccgtgaagctcctaaggtctcttggacaataaattctggygctct	2327
Q	2526	TGGATAGGCGCTAGACCTTCTCTTTGTTGTTTCTTCCAAACCCACACTCTCTTCTCAAGACCT	2585
D	2328	agagctcggggagctctcgtctctatactcctccaaagaagacccgctctctccagygata	2387
Q	2586	CATTCCACAGATAAAGTGGAAATGGTTTATATGATACAAAATAGCTTCTTGAAAGCTC	2645
D	2388	tttccctctttaaagtctccaggcgtctacagccgccaacaaactttaaagagatg	2447

QY	2646	TAGGATGGCGCGGTGGTTTAGTATTGGGAAGGCCTCAGCTACCTCGATTCCGTGGGGTC	2705
Db	2448	cgcctgaagcccgipccttctgtatgataygaaacctgaigyaacgtcttcataccccttgcat	2507
QY	2706	GAAATTGCTGCAGGGGGATATCGAATAATTCCTACACCCTATGATCTCTAGACTTGTTCTGT	2765
Db	2508	tcggttgaaaaaaaattccgcgaagatgaaaaaaaaataattccggattctctcaggccaacat	2567
QY	2766	TTCCGATGCTATCGTTAACATCCCCCAATCTACAGGACTCTTGATGATGAGGCCAGATC	2823
Db	2568	tgyrgatgigtatcgttaanaaatcccgcttcgctaactctcaatgycatgcygagccctc	2627
QY	2826	TTGGAAAATTCCGGGTGCGCATCTTTCAAAGACAGCATTTTACTGAGGGGTAGCAACA	2883
Db	2628	ttagacttcgctatgtaanaaacctcgacagaagcctctttaagaaygctlgaaagcca	2687
QY	2886	CTACGCTTAACAACCTCCAATTGTAGCTCTTTCGGACATTACGCTATGGAACCTCCGTGATC	2945
Db	2688	tctgaactcttcacctcatgtagaacctctctggygaagctgttlatgactcttcygtctc	2747
QY	2946	TTTCAAGGAACTAACATGTAGATGTGTGTGACCAACTCCGATCTTAG	2991
Db	2748	agcacacatctacaatgylagtatgtyvggctaataatcattcatctcsg	2793

RESULT	8	
AAA30854		
ID	AAA30854	standard; DNA: 2790 BP.
XX		
AC	AAA30854;	
XX		
DT	29-AUG-2000	(first entry)
XX		
DE	Chlamydia antigen CPN100639	gene open reading frame.
XX		
KW	Chlamydia antigen; diagnosis: infection; community acquired pneumonia;	
KW	therapy; upper respiratory tract disease; bronchitis; sinusitis;	
KW	asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;	
XX	ds.	
OS	Chlamydia pneumoniae.	
XX		
FH	Location/Qualifiers	
FT	1..2790	
FT	CDS	
FT	/*tag= a	
FT	/transl_except= (pos:1639..1640; aa:Phe)	
FT	/transl_except= (pos:2613..2617; aa:Val)	
FT	/transl_except= (pos:2641..2643; aa:Cys)	
FT	/transl_except= (pos:2702..2704; aa:Pro)	
FT	/partial	
FT	/product= Chlamydia antigen CPN100639	
FT	/note= "no stop codon given"	
XX		
PN	WO200032794-A2.	
XX		
PD	08-JUN-2000.	
XX		
PF	01-DEC-1999;	99WO-CA01147.
XX		
PR	01-DEC-1998;	98US-0110339.
PR	01-DEC-1998;	98US-0110340.
PR	01-DEC-1998;	98US-0110427.
PR	01-DEC-1998;	98US-0110428.
PR	01-DEC-1998;	98US-0110438.
XX		
PA	(CONN-)	CONNAUGHT LAB LTD.
XX		
PI	Murdin AD,	Oomen RP, Wang J;
XX		
WP1	2000-412339/35.	
DR	P-PSDB; AA90240.	
XX		

AC AAX06823;  
 XX 26-APR-1999 (first entry)  
 XX Chlamydia pneumoniae surface exposed protein Omp11 DNA.  
 XX Omp11, outer membrane protein 11; surface exposed protein; antigen;  
 KM infection; diagnosis; vaccine; atherosclerosis; asthma; ss.  
 XX Chlamydia pneumoniae.  
 OS  
 XX WO9858953-A2.  
 PN  
 XX 30-DEC-1998.  
 PD  
 XX 19-JUN-1998; 98WO-DK00266.  
 PF  
 XX 23-JUN-1997; 97DK-0000744.  
 PR  
 XX (BIRK/) BIRKELUND S.  
 PA (CHR/) CHRISTIANSEN G.  
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;  
 PI Myind P;  
 XX WPI: 1999-105610/09.  
 DR P-PSDB: AAM88424.  
 PT Species-specific test for identifying mammals infected with  
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for  
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding  
 PT these proteins  
 PS Claim 6; Page 62-63; 115pp; English.  
 XX  
 CC This DNA sequence codes for the novel 97.6 kDa surface exposed  
 CC protein Omp11 (see AAM88424) of the human respiratory pathogen  
 CC Chlamydia pneumoniae. By generating antibodies against C.  
 CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)  
 CC was obtained which reacted with outer membrane proteins. The  
 CC antibody was used to identify the genes (see AAX06816-27) encoding  
 CC Omp4-Omp15 proteins (see AAM88417-28) in an expression library of  
 CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:  
 CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in  
 CC the other, and encode polypeptides of about 89.5-100.3 kDa and  
 CC about 56.1 kDa. The invention provides a new species specific test  
 CC for identifying mammals (including humans) infected with Chlamydia  
 CC pneumoniae. The test comprises detecting antibodies specific for  
 CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer  
 CC membrane proteins, especially by PCR. The proteins are also used  
 CC in the diagnosis of C. pneumoniae infection in mammals. The  
 CC nucleic acids and proteins can also be used in the immunization of  
 CC mammals, the nucleic acids being particularly useful as DNA  
 CC vaccines for effecting in vivo expression of antigens. The  
 CC vaccines may also prevent atherosclerosis and bronchial asthma,  
 CC which are possibly associated with C. pneumoniae.  
 XX  
 SQ Sequence 2793 BP; 803 A; 681 C; 567 G; 742 T; 0 other;

Query Match 17.2%; Score 551.2; DB 20; Length 2793;  
 Best Local Similarity 52.7%; Pred. No. 6,7e-129;  
 Matches 1416; Conservative 0; Mismatches 1213; Indels 57; Gaps 8;

Db 261 gacattactggaagagatctacttcttcaacacgtagatgcyggttcgaatgc 320  
 QY AGGTGCTGCTGCATCTACACAGCAATACATCTTCTTCAGGTTTCTTACT 569  
 Db 321 agagctgcggaagcaaacactgtctgataaagcccttaacatcagattctcaact 380  
 QY GAGTTTATTCCTCTCTACACAGCGTTACTACAGGTGAGGAAAGCTTTCCTGACG 629  
 Db 381 ttcttcatcgcagctccttgaactacagctgcttcaggaaaaagttacttaagttcgc 440  
 QY AGAGGCGTAATTTAGAAAAATTCGTAACTGTACTTCTGGGAATTTTCTACTGC 689  
 Db 441 agagcccttaactcttaacgataatgaaacgattctctttagcacaacgctccaatga 500  
 QY AG-----ATGCTGAGCTATCAAGAGCGCTTTCTTTTACTGCGACTTCTGG 740  
 Db 501 agctaatacaatgcygagcgatcaccaaaaactcttctatcttctggaatacctc 560  
 QY AGATGCTCTTTTGTATCACTCTTCATCAACAAGGAGGAGCAATGCTACTACAGC 800  
 Db 561 ttctataaccttcactagtaataagcaaaaaaataagtgagcgactataagctctgc 620  
 QY AGCGCTCGCATATGCAATTAACACAGGTTATGTATCTCTATCTACATAGCGTCTAC 860  
 Db 621 ggcgtcaagatattcaagaaacacgcgcagtgtagtcttattgataataaaggaagaac 680  
 QY GTCAAGAGCGCCCTATCGATGATGAAGCGACGTCGATCTATGCAACAACAATTTCTATA 920  
 Db 681 tgggagcgcggtcctcggtcttggagcagctcctcgattcctcaaaaatagctccctt 740  
 QY TTTTGAAGGAATGCAATGCAAAACTACT-----GGCGGTGATCTGCAACACCA 971  
 Db 741 ctctctggaacaactgcaacaagatgcgcagcaagggcgggccatttttgaaaa 800  
 QY GGCAGTGTGATCTCTGATGATATATCTTCAACATTAAGACTGTGATTTGCTTCAA 1031  
 Db 801 aacagggagactcctactacttacttacttctggaataaagcttgcacttcgcgagaa 860  
 QY CGTAGCAAAACAAGCGGTGGCGCATTCATGCTAAAGACTAGCCCTTCTCTGAGG 1091  
 Db 861 ctctcttaactaaagcggaagcatctgcccattgtagacttcttcggtcgtcg 920  
 QY CTTTACAGAGTTTGTACCAATTAATGTCTCATCAGCAACTCT--AAGGGGGTGTAT 1148  
 Db 921 cccataacctattcaataataatagatgcggaaacacgctgcaggcaagggcggtat 980  
 QY CAGCATCGATGCTTCAGAGAGCTCAGTCTTTCGACAGAGACAGAAACATTAACCTTTGT 1208  
 Db 981 tgaattgcgagctcgtgattcttaagctctcctcgaacaatcaaggagacatcagcttcc 1040  
 QY AAGAAATACCTTACACACACCGGAAGTACGATACCTTAAAGTAAATGCGATCAACAT 1268  
 Db 1041 tggaaacacttaacacttaaccc---tccgcgcacaactcgacaggaatgattctacc 1097  
 QY AGGAAGTAAAGCGAAATTCACAGGAATTAAGGGCTGTCTAAATAATCAATCAATTTCTCTA 1328  
 Db 1098 gggatcgtcagcaaaaataaactaaagcaacttaagggcccaagccaactatctcttcta 1157  
 QY TGATCCCATCACTTCAGAAAGAAC---CTCATCAGAGCTATTAAGATTAATAACGGCTC 1385  
 Db 1158 tgaatcgatgcatcttaacacacagagctcagagcttgcacatcaacaaacgga 1217  
 QY TGGCGGAGCTTCAATCATATCAAGAGACATTCATTTCTGAGAAACCTCAACAGC 1445  
 Db 1218 tagcaactcgcctttagattatcaagaaagattgatttcttgggaaagaaagcctctcgc 1277  
 QY AGATGAATTAAGTGTGTGCAATTAATAATCTTCAATTCAGAGCGACGCTCCATATC 1505  
 Db 1278 agatgaagcgaagacgtcgtataactcaactatataaagaacaacatggtctcagc 1337  
 QY CGGAGGAAGTTATGTACTCAAAAAGGAGTCACTTAAAGAGACGAGACTTCTCTCAAGA 1565  
 Db 1506 tttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt



QY 2603 TCGAATGGTTATATACAAATAGCTTCTGAAAGCTCTAGTAGGCGGTGTT 2662  
 DB 2399 TCGAATGGTTATATACAAATAGCTTCTGAAAGCTCTAGTAGGCGGTGTT 2458  
 QY 2663 TTAGTATGGAGGCTCTTAACTCTGATCTCTGAGGCGCAAAATTCGTGCAAGG 2722  
 DB 2459 TTAGTATGGAGGCTCTTAACTCTGATCTCTGAGGCGCAAAATTCGTGCAAGG 2518  
 QY 2723 ATATGAGAGATCTCTAACCCTATATCTCTGAGATCTTCTGATCTCTGATCT 2782  
 DB 2519 ATATGAGAGATCTCTAACCCTATATCTCTGAGATCTTCTGATCTCTGATCT 2578  
 QY 2783 ACATGCTCCATCTACAGCGCTCTGAGATCTCTGAGATCTCTGAGATCTCTGAG 2842  
 DB 2579 ACATGCTCCATCTACAGCGCTCTGAGATCTCTGAGATCTCTGAGATCTCTGAG 2638  
 QY 2843 GCAATCTTTTCAAGACGAGCTTTTCTGAGGCTGAGCAACACTAGCTTCAACTCCA 2902  
 DB 2639 GCAATCTTTTCAAGACGAGCTTTTCTGAGGCTGAGCAACACTAGCTTCAACTCCA 2698  
 QY 2903 ATGTGAGCTCTTGGGACATTACGCTATGAGACCTCGTGGATCTTCAAGAACTACATG 2962  
 DB 2699 ATGTGAGCTCTTGGGACATTACGCTATGAGACCTCGTGGATCTTCAAGAACTACATG 2758  
 QY 2963 TAGATGTTGGTACCAACTCGATTC 2988  
 DB 2759 TAGATGTTGGTACCAACTCGATTC 2784  
 RESULT 6  
 AAA30853  
 ID AAA30853 standard; DNA: 3000 BP.  
 AC AAA30853:  
 DT 29-AUG-2000 (first entry)  
 XX Chlamydia antigen CPN100639 full length coding sequence.  
 DE Chlamydia antigen CPN100639 full length coding sequence.  
 XX Chlamydia antigen; diagnosis: infection; community acquired pneumonia;  
 KW therapy; upper respiratory tract disease; bronchitis; sinusitis;  
 KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;  
 KM ds.  
 XX Chlamydia pneumoniae.  
 OS  
 XX Key Location/Qualifiers  
 FH 101..2893  
 FT /tag- a  
 FT /product- Chlamydia antigen CPN100639  
 CDS  
 XX MO200032794-A2.  
 XX 08-JUN-2000.  
 PD  
 XX 01-DEC-1999; 99MO-CA01147.  
 PF  
 XX 01-DEC-1998; 98US-0110339.  
 PR 01-DEC-1998; 98US-0110340.  
 PR 01-DEC-1998; 98US-0110427.  
 PR 01-DEC-1998; 98US-0110428.  
 PR 01-DEC-1998; 98US-0110438.  
 XX  
 PA (CONN-) CONNUGHT LAB LTD.  
 XX  
 PI Murdin AD, Oomen RP, Wang J;  
 XX  
 DR WPI: 2000-412339/35.  
 DR P-PSDB: AAY90240.  
 XX  
 PT Nucleic acids encoding polypeptide antigens from Chlamydia useful for  
 PT preventing, diagnosing and treating diseases such as community acquired

PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset  
 PT asthma -  
 XX  
 PS Claim 2; Fig 7; 17app; English.  
 XX  
 This sequence encodes a Chlamydia antigen of the invention, designated  
 CC CPN100639. The nucleic acids (and their complementary sequences) may be  
 CC used as diagnostic agents for detecting the presence of nucleic acids  
 CC encoding Chlamydia antigens in samples according to standard methods,  
 CC and therefore, for diagnosing Chlamydia infections. For example, they may  
 CC be used as primers and probes for diagnostic polymerase chain reaction  
 CC (PCR) assays. Antisense sequences may be used to down regulate  
 CC expression of the proteins and may be used to treat infections. The  
 CC nucleic acids may also be used to produce the protein antigens they  
 CC encode according to standard recombinant DNA methodologies. The  
 CC proteins may then be used as antigens for the production of antibodies  
 CC (i.e. as vaccines) for preventing infection by Chlamydia. The  
 CC antibodies may also be used as diagnostic reagents for detecting  
 CC infections. Chlamydia is a pathogen implicated in the development of  
 CC (for example) community acquired pneumonia, upper respiratory tract  
 CC disease (especially bronchitis and sinusitis) asthmatic bronchitis,  
 CC adult-onset asthma and acute exacerbations of asthma in adults.  
 XX  
 SQ Sequence 3000 BP; 858 A; 726 C; 592 G; 824 T; 0 other;  
 Query Match 17.38; Score 555; DB 21; Length 3000;  
 Best Local Similarity 52.68; Pred. No. 7.6e-130;  
 Matches 1430; Conservative 0; Mismatches 1230; Indels 57; Gaps 8;  
 QY 330 AACGTTACTCTCAAAATCTCAGCCACACATATCTCTAACAAGAGATGCTTTT 389  
 DB 241 AATCTCAGCAGATGCTCAATGAGCAATGATGCTCTCAGCAATGCTCTT 300  
 QY 390 CGAGCTGGAAGAGGACCTCCCTTATCTGACAGTTGTTTAAAGCAACACGCAATCT 449  
 DB 301 CGATGCTGGAAGAGGACCTCCCTTATCTGACAGTTGTTTAAAGCAACACGCAATCT 360  
 QY 450 TACCTTCTGGGAGAGCTGATACCTTACGTTGGCTTATACATCTGCACTATGC 509  
 DB 361 GACATCTAGTGAAGAGGATCTCATTACCTTACGTTGGCTTATACATCTGCACTATGC 420  
 QY 510 AGGTGCTGCTGATCTCAACAGCAACGCAATTAAGAACTTCTCAGAGTTTCTTACT 569  
 DB 421 AGGTGCTGCTGATCTCAACAGCAACGCAATTAAGAACTTCTCAGAGTTTCTTACT 480  
 QY 570 GAGTTTGATCTCTCTAGCAGACGCTTACTACAGCTCAGGAAAGCTTCTCAGC 629  
 DB 481 TCTCTCAGTCAAGCTCTGAGCAATGCTTCTCAGGAAAGCTTCTCAGC 540  
 QY 630 AGGAGGCTTAATTAAGAAATATGTAACCTGTAAGTGTGCTGCAATTTTCTACTGC 689  
 DB 541 AGGAGGCTTAATTAAGAAATATGTAACCTGTAAGTGTGCTGCAATTTTCTACTGC 600  
 QY 690 AG-----ATGCTGAGCTATCAAGAGGCTCTTCTTTTAACTGCACTTCTG 740  
 DB 601 AGCTAATAAATGAGGCTGATCAACAGCAATTAAGAACTTCTTCTGCACTTCTG 660  
 QY 741 AGATGCTCTTTTAAAGCAACCTTCTGATCAACAGGAGGAGCAATTTCTACTACAG 800  
 DB 661 TCTCTAATCTCAGTCAATTAAGCAACCTTCTGATCAACAGGAGGAGCAATTTCTACTACAG 720  
 QY 801 AGGAGCTGCTGATCAACAGCAATTAAGCAACCTTCTGATCAACAGGAGGAGCAATTTCTACTACAG 860  
 DB 721 GGTCTGAGGCTGATCAACAGCAATTAAGCAACCTTCTGATCAACAGGAGGAGCAATTTCTACTACAG 780  
 QY 861 GTGAGGAGGCTGATCAACAGCAATTAAGCAACCTTCTGATCAACAGGAGGAGCAATTTCTACTACAG 920  
 DB 781 TGGGAGGAGGCTGATCAACAGCAATTAAGCAACCTTCTGATCAACAGGAGGAGCAATTTCTACTACAG 840  
 QY 921 TTTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 971  
 DB 841 CTCTCTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900

181 ttttaagagccctggaaaagcaacctccctatctgacagctgttttaagaaacacagagc 240  
445 AATCTTACTCTTCTGGGGAACGGTCTACTTAACTTGGCTTTATAGATGCTGGCACT 504  
241 aactctaccctcttgggaaagcgtcaatgacttaacgcttgccttataagatgctggcact 300  
505 CATGCAGTGTGCTGCTCACTCACTCAACAGCAATAAAGATCTTACCTTCTCAGAGGTTTTC 564  
301 catgcaaggtgctgctgcacatctacaacagaataaagaaacttaacctctccaggttttcc 360  
565 TTACTGAGTTTGAATTCCTCTCTAGCACAAACGGTTTACTACAGGTTCAGGAAAGCTTTTC 624  
361 ttaactgagtttgatctccctccctcctagcacacagcttactacagtgcaaggaagcttcc 420  
625 TCAGCAGAGGCGGTAAATTTAGAAAAATATCTTAACCTGTAGTTCCTGGAAATTTTCT 684  
421 tcaagcagagggcgttaaatctagaataatctglaaactgtagctgctgggaattttc- 479  
685 ACTGCAGATGTGAGAGCTATCAAGAGAGCGTTCCTTTTAACTGGCACTTCTGGAGAT 744  
480 actgcagaatggtggagcctatacaaggagcgtcttcc- ttaacgycacctctcggagat 538  
745 GCTCTTTTAAAGAACTCTTCAACAAAGGAGGAGCAATTCCTACTACAGCAGGC 804  
539 gctcttttagtaacaactctcctcatcaacaaggagagcaattgctactacagcagcg 598  
805 GCTCCGATAGCAATTAACACAGGTTATGTATTCTTCTTAACTTACGCTACGCTAGTGA 864  
599 gctcgataagcaataaacaaggtatgttagatctcctatacaacaaagcgtctcagta 658  
865 GGAGGCGCTATGATGATGAAGAGCAGTGCATCTATCGAACACAAATTTTATATTTT 924  
659 ggaaggcgctatcgatgataaggaagcagctgctatctatcgaaacaatcttataattt 718  
925 GAAGGAATGCAAGCAAACTACTGGCGGTGCGATCTGCAACCAAGCGAGTGATCT 984  
719 gaaggaaatgcagcgaaactactcgtgagctgcatcgacaaccaagcgagtgatct 778  
985 CCGTAACCTATATCTCTAACAATAAGACTGATCTTGGTCAAAAGCTACAGCAACA 1044  
779 cctggaactgataatctctcaacaataagactcgtacctgttctcaaaagcgaagaataa 838  
1045 AGCGGTGCGCCATCCATCCTATAAAAGCTAGCCCTTCTCTGAGGCTTTACAGAGTT 1104  
839 agcggtggcgccatcocalgctaaagaagcctagcccttctctcggagccttaagaagtt 898  
1105 CTACGAATTAATGTCTCATCAGCAACTCCTAAGGGGGGTCTATCAGCATGCTCTCA 1164  
899 ctacgaaataatgctcctcaacagcaactcctaaggggtgctatacagcatgagctca 958  
1165 GGGAGACTAGTCTTTCGACAGAGCAGGAAACATTAACCTTTGTAGAATAACCCCTTACA 1224  
959 ggaagctcagctctctcgcagagagaagaacaataccttctgtaagaataacccctta 1018  
1225 ACAACGGAAGTACCGATCTACTCTAAACCTAATGGATCAACATAGGAAGTAACGGGAAA 1284  
1019 acaaccggaagtaacgatactactcctaaagtaatgcatataacataggaagtaacgggaaa 1078  
1285 TTCACGGAATTAACGGGCTCTTAAAAATCATACAAATTTCTTGTATGATCCCATCTTCA 1344  
1079 ttcaaggaattcagggctgctaaataatacaaatcttctatgataccctca 1138  
1345 GAAGGAACCTCATCAGACCTATTTGAAGATAAATAAGCGGTTCGGGAGCTCTCATCA 1404  
1139 gaaggaacctcatcagagatctgaagataaataaagctcgcggggaagctctcaatcca 1198  
1405 TATCAAGAAAGATTTCTATTTTCTGAGAAACCTTAACAGCAGATGAACCTTAAAGTTGCT 1464  
1199 tatcaaggaagcatctctatctctcgtgagaacccctaaacagcagatgaacttaagtgct 1258  
1465 GACAATTTAAATCTTCAATTCACGACGAGCTCTCCCTATCCGGAGGAAAGTTATGCTTA 1524

1259 gacaattaaactctcatcctcagcagccagctctccctatcccgaggaagtaattgcta 1318  
1525 CAAAAGGAGTCACTTTTAGAGAGCAAGAGCTTCTCMAAGAGAGCGCGTTCCTCCGCGC 1584  
1319 caaaaggagctacttaagagcagagctctctccaagaggcggtctctctccgcgc 1378  
1585 ATGATTCAGAGACGACATTAATCAACTACAGCTGGAGAGTATTACAAATCAGAACTAGGA 1644  
1379 atgagatcaggaagacatctatcaactacagctgggagatatttaacatcaacgaaactaga 1438  
1645 ATCAATGTTGACTCTTACGCTTTAAGCAGCCCGCTACGCTTAACAGCAAAAGTGTCTTA 1704  
1439 atcaatgtagactcttgagcttgaagcagcccgctcaagcttaacagcaaaagtgcttca 1498  
1705 AATTAAGTATGCTATCTGAGAGCTCAACGATGATTGAAGGGAACCTTATGAA 1764  
1499 aactaaagatcgtatcctgggaagcccaaccgatatgatatgaaaggaacttaatgaa 1558  
1765 ACTCATGTTTCAGCCATGACAGCAGCTCTCTCTATTTAAAAATCAGCGTTGATGCTGAT 1824  
1559 agctatagctcagccatcagcagcctctctctctatataaatacaagcttgatgctgat 1618  
1825 GTTGATTAATCAAGCTTGACATCAAGACCTTATCCCTGTTCTGCTGAGATCTTAATTA 1884  
1619 gtgataactaaagttgacatcagcagccttaaccctgtctcgtcggagatcctaattca 1678  
1885 GAATCGGATTCGAAGGACAAATGSAATGTTAATGAGTACAGGATACACTCAATACA 1944  
1679 gaataagatcccaaggaaatgaaatgtaattgaaactaagatacagctcaataataa 1738  
1945 AAAAGGCGCACGGCAACTTGGACCAAAACAGAGTTTGTCCAGCCCGCAAGAAAATCT 2004  
1739 aaaggagccagcgcaacttggacccaacagatattgttccagcccggaagaataatc 1798  
2005 GCGTTAGTATGCAATTAACCTATGGGAGCTTTACTGATTCGCTCTGTGAACAGCT 2064  
1799 ggcattagatgtaaatccctatggggagcttcttaactgaaatcgcctctgcaaacgct 1858  
2065 GTAGAGATGCGGCAACTGATGMAACAACAAGAGTTTGTGGGTTTCCTGCATGAGC 2124  
1859 gtaagagatcggcgcaactggtatgtaaacacaacaaagttcttggttccctccatgag 1918  
2125 AACTTCCTGCATTAAGACTGAGATGAAATGCAAAAGCTTCCTGCTATACCTCTGAGGC 2184  
1919 aacttcctgcatlaagactgagatgaaatcgaaaggtcttcgctataccctcgtagagc 1978  
2185 TACGTCATGGGTGGAAGTCTCACTCTCTAAGACGACTTTTATACCTTTCGCTTCGC 2244  
1979 tactgatacgtggaagtgctcacactcctlaaagacgacctatcttaacttgctgctgc 2038  
2245 CATCTCTTGTGAGAGACAAAGATTTT - TATCGCTCAACAACACTGTAGAACCTAGC 2302  
2039 catctcttctgctagagacaagaattgtttatatatgcttcaacaacatctagaactaag 2098  
2303 GTGGAACCTTATTTCTCAAGCACTCTCATACCTTACACCCCAAAACTATTTGAGATTAG 2362  
2099 gtggaacttattcttcaagcaactctcataccctcaacccaacataattgagttag 2158  
2363 GAAGGCAAAAGTTTCTGAAATCAGCATATGAAAAATTCCTTAGGGAATAATCCCTAGCCT 2422  
2159 gaaggaacaagttcttctgaaatcagcatatagaanaattcccttgaggaatctccctagct 2218  
2423 TCGATGTCGAATTTCTGTCAGCACTTCAAGAACCGTATGGAAGCGCACTATACCTCAT 2482  
2219 tggatgccaagatctcgttccagccatcagacaacocgtatlgbaaacgactatactcat 2278  
2483 TCCCAGAATCCGAAGTTCTTGGAGCAACGAGGTATATAGCTGTGGTATTCGCCCTAGACC 2542  
2279 tgcagaaatccgaagttctcttgagcaacagtgataagcttggtatcgcgcctagagc 2338  
2543 TTCCTTTTGTCTTTCGAACCAACATCTCTTTCAAGCCTTCAATTCACAGATGAAG 2602  
2339 ttccttltgtcttcccaaccacatcccttctcaagaccttcatccacagatgaaag 2398







OY	105	GATAAAAATTTCTTGACAGCGTGTTTGTGCATCTTTAACTTGATTTACTATTTTGTTC	164
Db	1	gataaaatctcttgaacgcgtctttcttcaactcttaacttgattactatttcttctcta	60
OY	165	TATTGATGGCAATGTTCCTATAAAAAAAGCATTTACCATGAAGACCTGGATTCCTTG	224
Db	61	tatgtagcgaaatagttctcttaaaaaacaagaacatctaacgaagaaactcgaatccttg	120
OY	225	GGTTTATGTTTCCGCGGTGTAGCTTTTCATGTCCACCTACACGTACATAGCTAAAGAGA	284
Db	121	ggttttagttcttcctcogtlttagcttcttcacgtccacacgtcaactagctaacgagga	180
OY	285	ACTTTTATACCTGTGATGATAGCTTTAATGGAATATCGATTGAGAACGTTTACTCCAA	344
Db	181	acttttaacctgatgatagctttaatgaaatcgtcttcaggaaacgtttactccaa	240
OY	345	AACTTTCAGCCACAACATATTCCTTAACAGAGATGTCTTTTACGAGCCTGGAAAAAG	404
Db	241	aacttcagccaacaacatctcttaacacaggaatgtctcttcttcagacccgagaaag	300
OY	405	CACTCCTTATCTACAGTTGTTTTAAAGAAACACAGCAATCTTACCTTTGGGAA	464
Db	301	caatcccttatctgaacgattgttttaagcaaacacaaagacaactctactctcttgggaa	360
OY	465	CGGTCAATAGCTTAACTGTTGGCTTTATAGATGCTGGCACTATGCAAGTGCTGTCA	524
Db	361	cggtcatagcttaacgittgttgccttataagatgtgcactcatgcaagtgctgcac	420
OY	525	TACAAACGCAAAATAGAAATCTTACCTTTCAGGGTTTCCCTACAGATTTGATTCCTC	584
Db	421	tacaacagcaataaagaatcttaacctctcaaggttcttcccttaacgaatttgaatcc	480
OY	585	TCCTAGCACAAACGGTTACTACAGTCCAGGGAACGCTTTTCCACAGAGAGCGTAAAT	644
Db	481	tccatagaacaacagttactacacagtgcaaggaaagcttctcgaagaagtgataaatt	540
OY	645	AGAAATATTTCTGAACCTGTAGTGTGTGGGAATTTTCTACTGCAGATGTGAGCTAT	704
Db	541	agaaatatctgtaaaccttgtagtgcggaatttcttactgagaatgtygagcat	600
OY	705	CAAGGAGCGCTTCCCTTTTAACCTGGCACTTCTGGAGATGCTCTTTTAGTAACAATC	764
Db	601	caaaagagcgcttctcttcttaacatgcacatcttgagatgctctttagtaaaactc	660
OY	765	TTTCATCAACAAAGGAGAGACCAATTTGCTACTACAGAGCGCTCCGATAGCAAAATPAC	824
Db	661	tttcatcaacaagaagaggaacatgtctactacagcagcgctgcataagcaataaac	720
OY	825	AGGTTATGTTAGATTCCCTATCTAATACATAGCGTCTACGTCAGAGAGCGCTATCGATGA	884
Db	721	agttatgttagattctctactactaactagtgctactgcataagagcgctatgatgata	780
OY	885	AGGCACGTCCGATACTATTCGAACAACAATTTCTATTTTGAAGGAATGCAGCGAAAC	944
Db	781	aggaacgtgatatactatcgaacaacaattctataatttgaaggaatgacgcyaaac	840
OY	945	TATCGGGGTCGATCTGCACACCAAGCGAGATGATCTCTGAAACGATATCTTCAA	1004
Db	841	tactggcggtgtgatctgcacaacacaagcgatgatctcttgaaactgataatctccaa	900
OY	1005	CAATTAAGACTCTGATCTTTGCTTCAACAGCTAGCAGAAACAGCGGTGGCGCATCATGC	1064
Db	901	caataagactctgatcttcttgcttcaaacgtlagcagaacaacgctgtgcacatccatgc	960
OY	1065	TAAAAAGCTAGCCCTTTCCTCTGGAGGCTTTACAGAGTTTCTACGAATTAATGTCTATC	1124
Db	961	taaaaaactagcccttctcctcttgaggtcttcaagagtttctacagaataatgtctcatc	1020
OY	1125	AGCAACTCTTAAGGGGGGGCTCTACGACATCGATGCTCTCAGAGAGAGCTCAGCTTTCGC	1184
Db	1021	agcaactcttaagggggtgtctacacgaatcgatgctctcctcaggaagctcgaagcttctgc	1080
OY	1185	AGAGACAGGAAACATTACCTTTTGAAGAAATACCTTACACACCGGAATACCGATAC	1244

Db	1081	agagacaggaaaccttacccttctgtaagaatatacccttacaacaacccggaaatgacgataac	1140
Qy	1245	TCCTAAACGTAAATCGATCAACATATAGAACTAACGGGAAATTCACGGAAATTCGGGCTGC	1304
Db	1141	tcccaaacgtaatgcatcaacaataagaagtaacgaggaaatccaaggaattacgagctacgctgc	1200
Qy	1305	TAAAAATCATCAAAATTTTCTTATATGATCCCATCTCTAGAAAGAACTCATCAAGAGCT	1364
Db	1201	taaaatatacaattcttcttctatgtatcccatcaacttcagaaggaacttcacagact	1260
Qy	1365	ATTGAAGTAAATTAACGGCTCTGCGGAGAGCTCTCAATCCATATCAAGGACAGATCTTAT	1422
Db	1261	attgaaataataaagcgtctcgaggagcttccaatccataccaaggaagcattcat	1320
Qy	1425	TTCTGAGCAAAACCCCTAACAGAGATGAACCTTAAAGTTGGGACATTTTAAATCTTCAT	1488
Db	1321	tctcgagaaaccctcaacagaagatgcaacttaagaattgctgcacaatttaaatcttcat	1380
Qy	1485	CACGCAGCCAGTCTCCCTATTCGAGAGAAAGTTATTTGCTACAAAAGGAGTCATTTTAGA	1544
Db	1381	caaggaagcagcttccctatactcgagaagattatgctacaagaagggtcaactttaga	1441
Qy	1545	GAGCAGAGCTTCTCTCAAGAGCGCGTTCTCTCCTCGCGATGGATTCCAGGAACACAT	1600
Db	1441	gagcagagcttctctcaagaagcggttctctctcctcgatgattcaggaaagcatt	1500
Qy	1605	ATCAACTCTACACTGGAGATTTTAAATCACGAACCTAGAAATCAATCAATGACTCCCTTAG	1664
Db	1501	atcaactcagcgtcgaggatataaacaatcaagaaccaggaaatcaatgattactcttagg	1560
Qy	1665	TCTTAAGCAGCCCGTCAGCCTTAACAGCAAAAGTGCTTCAATTAAGATGATGTATCTGG	1722
Db	1561	tcttaagagcccgctcagccttaacagaagaagtgtctcaataaagtatgcgtatctagg	1620
Qy	1725	GAAGCTCAACTGATGTATATGAAGGAACTTTATGAAAGTCATATGTTTCAGGCATAGA	1788
Db	1621	gaagctcaacctgtatgatatgaaggaaacattatgaagatcatagtltcaagcaatga	1680
Qy	1785	CCAGCTCTTCTCTCTATTTAAATAACAGGTTGATGTCGATGTGATCTTAACGTTGACAT	1844
Db	1681	ccaagctctctctataataaatacaagcgtltgtagtgcgtatgtagtaactcaagaagttgaact	1744
Qy	1845	CAGCAGCCTTATCCCTGTTCCGTGCTGAGATCTTAATTCAGAAATACGATTCCTCAAGACA	1904
Db	1741	cagcagccttataccctgttctctgttgtaggttccataatcaagaatacgaattcccaaggaca	1800
Qy	1905	ATGGAATGTAAATTGAGCTACGATACAGTACAAATACAAAAGAGCCACGGCAACTTG	1966
Db	1801	atgaaatgttaattgagctaaagatacaagatacaatacaaaagagccacagcgaacttgg	1860
Qy	1965	GACCAAAACAGAGATTTGTGTCACACCCCGGAAACAAAATATCGCTTGTATGGAATATACCT	2022
Db	1861	gaccaaaacagatctgttcccaacccccgaagaanaatctgcgtatgataatgaatacccc	1920
Qy	2025	ATGGGAGCTCTTTACTGTACATTCGCTCTCTGCAACAAGCTGTATAGAGATGGCGCAACTGG	2084
Db	1921	atgggagatcttactgaatctgcctctcgcgaacgctgttagagatactgggcaacttgg	1980
Qy	2085	TATGGACACAACAAGGTTTCTGEGTTTCTCTCATATGACGAACCTCTCTCATTAAGACTGG	2144
Db	1981	tatgaaacaacaagaagttctctggttctctccatgaacgaacttctcgtataagaacttgg	2040
Qy	2145	AGATCAAAATGCGAAAGGTTTCGCTCATCTACTCTGGAGGCTATGTCATGCGGGAAGTGC	2204
Db	2041	agatgaaataatgcgaaggcttccgtctacactctggagtgcaatcogtctggaagttgc	2100
Qy	2205	TCACACTCTCTAAACACAGCACTATTACTCTTGGGTTCTGTGCACATCTTTTGTCTAGAGACA	2264
Db	2101	tcaactctctaaagaagcactattactcttgggtctcgtgcactctcttcttcaggagaca	2160
Qy	2265	AGATTGTTTATCGCTCAACAACAATCTAGAACCTAAGGTGAATTTATTTCTTCAAGCA	2324

QY	2161	GGTTTCGCTCACTACCTCTGGAGGGCTACGTCAATCGCTGGAAAGCGTCCACTCCTAAAGAC	2220
Db	520319	ggcttcgcgtcaaacctcggagagctacgctacgttcggtygaagtgtctccagctcccaagac	520378
QY	2221	GACCTATTTCACCTTTGGCGTCTGCACATCTCTTTCCTCTAGAGACCAAGAATTTCTTTATCGCT	2280
Db	520379	gacctattacctcttgctcgtctgcacatctctcttgctagagacaaagtgttttatcgct	520438
QY	2281	CACAACAACCTCTAGAACCTACGGTGGAACTTATTCTTCAAGCACTCTCATACCTTCMA	2340
Db	520439	cacaacaacctagaaacctaaacggtyggaacttattcttccaagcactctcaacctcaaa	520498
QY	2341	CCCCAAAACATTTTGAGATTAGGAAGAGCAAAAGTTTCTGATACGTATAGAAAAATTC	2400
Db	520499	ccccaaactattttagtgatgtaggaaggaagcaaaagtcttcgaatcagctatagaaaaatlc	520558
QY	2401	CCCTAGGAAATTCCTCCTAGCCCTTGATGTCACAGTTCCTTCACGCCATTCACACACCGT	2460
Db	520559	ccctagggaattccctccttagcctcttgatggtccaaagtttcgltccagccatcagaacaacglt	520618
QY	2461	ATGGAAAGCACTTACCTCCTCATCTTGCAGAAATCCGAAGTTCCTTGAGACCAAGGTGATA	2520
Db	520619	atggaaagcacttaaacctcaatcgtccgaaacccgaagttctcttgagcaacagagtgata	520678
QY	2521	GCTGGTGATACGGCGCTAGACCTTCCTTTGTTGTTTTCACACCACATCCCTCTTTCAAG	2580
Db	520679	gctggtggtatcgcgcgttagacctctctcttgtctctccaaaccacatcctctttccaag	520738
QY	2581	ACCTTTCATTCACAGATGAAAGTGGAAATGGTTTATGTATCACAAATAGCTTCTTGAA	2640
Db	520739	acctcttctccacagatgaaagtccgaatggtttatglatcacaanaataagctctctcgaa	520798
QY	2641	AGCTCTACTGATGGCGCGGTTCATTAGATTGGGAAGGGCGGTAACTCTGTGATCCGTG	2700
Db	520799	agctctagtgtgcccgtggtgttttagtatgtggaagcgtcttaaccctcctgatctctgtg	520858
QY	2701	GGTGGAAATTCGTGACAGGGGGATATCGAGATTCCTACACCTATGATCTCTCAGATTTC	2760
Db	520859	ggtcgaaatctcgcgcaggggagatalcgagatcttccatacctaigtctctccaaggttc	520918
QY	2761	TTTGTTTCCGATGTCATGCTGATAAATCCCAATCTACAGCACTCTGTGATGAGGCCA	2820
Db	520919	tttgttcttcgcatgctcatcgtaaaacatccccaactcagaagcctcttgtagaagccca	520978
QY	2821	GACCTCTGGAAATTCGCGGTGCGGAATCTTTCAGAGACGCAATTTTATCTAGAGGGGTAC	2880
Db	520979	gacctctggaataattcggcggtyggaactcttccaagacagcatctttactgagggttagc	521038
QY	2881	AACAACATACGCTACACTCCAAATTGAGACTCTTGGGAATTAACGCTATNGAACTCCGT	2940
Db	521039	aacaactacgctactacaactccaattgtagactctctcggacattacgctatgtagaactcgt	521098
QY	2941	GGATCTTCAAGGACATCAATATGTGATGTTGTCACAAATCCGATTCGTGATGTGCTAAA	3000
Db	521099	ggaatcttcaaggaactacaatgtatgtagtgtgtacccaactccgatctctatgtctaaa	521158
QY	3001	ACTCCCTAGTTCCTCTAGGAGTTCCTCATCTCTTTTATGGGAAATATTTCCTATAGGGAA	3060
Db	521159	actccctagttctctcttagagggtttctctcactctttaggaaataattctctataggaa	521218
QY	3061	TGCTTTCCTTGCAACGTGTAATAAATATACATTTGTCCCTCTCAAAAACATTTCTTTTA	3120
Db	521219	tgcttctcttgcaaacgtgtaaaaaataaacatctgtccctctcaaaaaaagatctcttta	521278
QY	3121	ATAATTTCTAGTATTAATTTTATTTTAAAAACAGTTAAATATTAATATAGCAATAATCTTA	3180
Db	521279	ataatctctagtataaatttattattttaaaaaacgtaataataataatagcaataatctta	521338
QY	3181	TTCTTATTGACTCTTTTTTTT 3200	
Db	521339	ttctctattgactcttcttttt 521358	

XX	RESULT# 4
XX	ID AAA30847 standard; DNA; 2950 BP.
XX	AC AAA30847;
XX	DT 29-AUG-2000 (first entry)
XX	Chlamydia antigen CPN100634 full length coding sequence.
DE	Chlamydia antigen; diagnosis; infection; community acquired pneumonia;
KW	therapy; upper respiratory tract disease; bronchitis; sinusitis;
KW	asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;
KW	ds.
XX	Chlamydia pneumoniae.
OS	
XX	
FH	Key Location/Qualifiers
FT	CDS 101..2887
FT	/tag=a
FT	/product= Chlamydia antigen CPN100634
PN	WO200032794-A2.
XX	
PD	08-JUN-2000.
XX	
PF	01-DEC-1999; 99WO-COA01147.
XX	
PR	01-DEC-1998; 98US-0110339.
PR	01-DEC-1998; 98US-0110340.
PR	01-DEC-1998; 98US-0110427.
PR	01-DEC-1998; 98US-0110428.
PR	01-DEC-1998; 98US-0110438.
PA	(CONN-) CONNAUGHT LAB LTD.
PI	Murdin AD, Oomen RP, Wang J:
DR	WPI: 2000-412339/35.
DR	P-PSDB; AAY90236.
PT	Nucleic acids encoding polypeptide antigens from Chlamydia useful for preventing, diagnosing and treating diseases such as community acquired pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset asthma -
PS	Claim 2; Fig 1; 174pp; English.
XX	This sequence encodes a Chlamydia antigen of the invention, designated CPN100634. The nucleic acids (and their complementary sequences) may be used as diagnostic agents for detecting the presence of nucleic acids encoding Chlamydia antigens in samples according to standard methods, and therefore, for diagnosing Chlamydia infections. For example, they may be used as primers and probes for diagnostic polymerase chain reaction (PCR) assays. Antisense sequences may be used to down regulate expression of the proteins and may be used to treat infections. The nucleic acids may also be used to produce the protein antigens they encode according to standard recombinant DNA methodologies. The proteins may then be used as antigens for the production of antibodies (i.e. as vaccines) for preventing infection by Chlamydia. The antibodies may also be used as diagnostic reagents for detecting infections. Chlamydia is a pathogen implicated in the development of (for example) community acquired pneumonia, upper respiratory tract disease (especially bronchitis and sinusitis), asthmatic bronchitis, adult-onset asthma and acute exacerbations of asthma in adults.
SQ	Sequence 2950 BP: 851 A; 670 C; 596 G; 833 T; 0 other:
XX	
XX	Query Match 92.2%; Score 2950; DB 21; Length 2950;
XX	Best Local Similarity 100.0%; Pred. No. 0;
XX	Matches 2950; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





hybridization-based assays or amplification-based assays. The protein may be used for diagnostic purposes, for their enzymatic or structural activity, or as a vaccine. The invention also describes (1) a probe comprising a hybridizing fragment of NI; (2) an isolated nucleic acid (N2) that hybridizes under stringent conditions to NI; (3) an expression cassette comprising NI under the transcriptional regulation of a transcriptional initiation region functional in an expression host, and a transcriptional termination region; (4) a cell comprising an expression cassette of (3) as part of an extrachromosomal element or integrated into the genome of a host cell as a result of induction of the expression cassette into the host cell, and the cellular progeny of the host cell; (5) a method for producing a plasmid comprising growing a cell of (4) where the protein is expressed and isolating the protein free of other proteins; (6) a purified polypeptide composition comprising at least 50 weight % of PI; and (7) a monoclonal antibody binding specifically to the peptide of (6).

Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

Query Match 100.0%; Score 3200; DB 21; Length 273254;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CAATGTCGAGAGACACTAACCCAGGAAATTCGATTTCATTAACCCACTTATATTA 60  
98235 CAATGTCGAGAGACACTAACCCAGGAAATTCGATTTCATTAACCCACTTATATTA 98176  
61 AATTCCTACCTGCGTCATTAATAATAGAAACTCAGAGAGTCAAGTAAATTTCTTGAC 120  
98175 AATTCCTACCTGCGTCATTAATAATAGAAACTCAGAGAGTCAAGTAAATTTCTTGAC 98116  
121 ACGTGTTCATCTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 180  
98115 ACGTGTTCATCTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 98056  
181 TCTCTAAAAAACAAGACATTAACATGAAGACTTCCTTGGGTTTAACTTCTCC 240  
98055 TCTCTAAAAAACAAGACATTAACATGAAGACTTCCTTGGGTTTAACTTCTCC 97996  
241 GTGTAGCTTCTCTACGTCACCTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 300  
97995 GTGTAGCTTCTCTACGTCACCTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 97936  
301 GATFACCTTAAATGAGAAATATCGATTACAGAACTTAACTTAACTTAACTTAACTTAACT 360  
97935 GATFACCTTAAATGAGAAATATCGATTACAGAACTTAACTTAACTTAACTTAACTTAACT 97876  
361 TATCTCTAACGAGAGATCTCTTCTTAAAGAGCTGAGAAAGGACCTCTTATCTGAC 420  
97875 TATCTCTAACGAGAGATCTCTTCTTAAAGAGCTGAGAAAGGACCTCTTATCTGAC 97816  
421 AGTCTTTTAAAGCAACACGAGCAATCTTACCTTCTTGGGAAAGGTCATAGCTTAAG 480  
97815 AGTCTTTTAAAGCAACACGAGCAATCTTACCTTCTTGGGAAAGGTCATAGCTTAAG 97756  
481 TTTGGCTTATAGATGCTGACACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540  
97755 TTTGGCTTATAGATGCTGACACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 97696  
541 AATCTTACCTTCTAGGCTTCTTCTTAACTGATTTTGAATCTCTCTCTAGCAACAGGTT 600  
97695 AATCTTACCTTCTAGGCTTCTTCTTAACTGATTTTGAATCTCTCTCTAGCAACAGGTT 97636  
601 ACTACAGGTCAGGAGAGCTTCTCTACAGAGAGGCGTAAATTTGAAATTTTGTATAA 660  
97635 ACTACAGGTCAGGAGAGCTTCTCTACAGAGAGGCGTAAATTTGAAATTTTGTATAA 97576  
661 CTGTAGTATGCTGGAAATTTTCTACGATGCTGAGAGCTATCAAAAGAGAGCGCTTTC 720  
97575 CTGTAGTATGCTGGAAATTTTCTACGATGCTGAGAGCTATCAAAAGAGAGCGCTTTC 97516  
721 CTTTAACTGACCTTCTGAGAGTCTCTTTTAACTAACACTTCTTCAATCAACAAAGGA 780

||||| 97515 CTTTAACTGACCTTCTGAGAGTCTCTTTTAACTAACACTTCTTCAATCAACAAAGGA 97456  
781 GGAGAAATTTGCTACAGAGCAGAGCGCTGCATAGCAAAATTAACAGGTTATGATTC 840  
97455 GGAGAAATTTGCTACAGAGCAGAGCGCTGCATAGCAAAATTAACAGGTTATGATTC 97396  
841 CTATCTAACATFAGCTCTACGTCAGAGAGCGCTATGATGATGAAGGACGCTGATCTA 900  
97395 CTATCTAACATFAGCTCTACGTCAGAGAGCGCTATGATGATGAAGGACGCTGATCTA 97336  
901 TCGAACCAAAATTTCTATATTTTGAAGGAAATGACAGCAAACTACTGCGGTGATC 960  
97335 TCGAACCAAAATTTCTATATTTTGAAGGAAATGACAGCAAACTACTGCGGTGATC 97276  
961 TCGAACCAAAAGGCGAGTCTCTGACATGATTAATCTTAACAAATTAAGCTGTATC 1020  
97275 TCGAACCAAAAGGCGAGTCTCTGACATGATTAATCTTAACAAATTAAGCTGTATC 97216  
1021 TTTGCTTCAAACTGACAGAAACAAGCGGTGCGCATCATGCTTAAAGCTAGCCCTT 1080  
97215 TTTGCTTCAAACTGACAGAAACAAGCGGTGCGCATCATGCTTAAAGCTAGCCCTT 97156  
1081 TCCCTGAGAGCTTTACAGAGTTTCTACGAATTAATGTCATGACCACTCTAAGGGG 1140  
97155 TCCCTGAGAGCTTTACAGAGTTTCTACGAATTAATGTCATGACCACTCTAAGGGG 97096  
1141 GGTGCTATCAGATGAGTCCCTCAGAGAGCTCATGCTTCTGAGAGACAGAAACAT 1200  
97095 GGTGCTATCAGATGAGTCCCTCAGAGAGCTCATGCTTCTGAGAGACAGAAACAT 97036  
1201 ACCTTGTATAGAAATACCTTACAAACCGGAGTACGATACCTTAAACGTAATGG 1260  
97035 ACCTTGTATAGAAATACCTTACAAACCGGAGTACGATACCTTAAACGTAATGG 96976  
1261 ATCAACATATGGAATTAACGGGAAATTCACGGAATTAACGGGCTGCTTAAATATCAATT 1320  
96975 ATCAACATATGGAATTAACGGGAAATTCACGGAATTAACGGGCTGCTTAAATATCAATT 96916  
1321 TTTCTTATGATCCCATCTCAGAGAGCTCATGAGAGCTTAAATTAATAC 1380  
96915 TTTCTTATGATCCCATCTCAGAGAGCTCATGAGAGCTTAAATTAATAC 96856  
1381 GGCCTGCGGAGCTCTCATATCATATCAAGAGAGATTTCTTCTGAGAAACCTTA 1440  
96855 GGCCTGCGGAGCTCTCATATCATATCAAGAGAGATTTCTTCTGAGAAACCTTA 96796  
1441 ACAGCATGAACTTAAGTCTGCAATTTAAATCTTCACTTACAGCAGCAGCTCTC 1500  
96795 ACAGCATGAACTTAAGTCTGCAATTTAAATCTTCACTTACAGCAGCAGCTCTC 96736  
1501 CTATCGGAGGAAATTTATGCTTACAAAGGAGTCACTTGAAGAGCAGAGCTTCTCT 1560  
96735 CTATCGGAGGAAATTTATGCTTACAAAGGAGTCACTTGAAGAGCAGAGCTTCTCT 96676  
1561 CAAGAGCGGCTCTCTCTCGCATGATTCAGAGACACATTAATCACTACAGCTGGG 1620  
96675 CAAGAGCGGCTCTCTCTCGCATGATTCAGAGACACATTAATCACTACAGCTGGG 96616  
1621 AGTATTAATACAGACCTAGGAATCAATGTTGACTCTTATGCTTAAAGAGCCGCTC 1680  
96615 AGTATTAATACAGACCTAGGAATCAATGTTGACTCTTATGCTTAAAGAGCCGCTC 96556  
1681 AGCTTACAGCAAAAGGCTCTCAATTAAGAGTGTATGAGAGGTCACACCTGAT 1740  
96555 AGCTTACAGCAAAAGGCTCTCAATTAAGAGTGTATGAGAGGTCACACCTGAT 96496  
1741 GATATTAAGGAGCAATTAATGAATGATATGATGATGATGATGATGATGATGATGATGAT 1800  
96495 GATATTAAGGAGCAATTAATGAATGATATGATGATGATGATGATGATGATGATGATGAT 96436  
1801 TTAATAATCAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860

QY	1621	AGTATTTACATTCACGAACCTTAGAATCATGTTGACTCCTTAGCTTTAAGACGCCGTC	1650
Db	1621	agttatlaacatcacgaacacttaggaatcaatgaatgctaccccttagtcttaagaacgcgctc	1680
QY	1681	AGCCCAACGAGAAAAGGTGCTCAAAATTAAGTATGCTCTCGGGAGCGTCAACCTGATT	1740
Db	1681	agccttaacagcaaaaaggctgctcaaatbaagtgatcgtatccggaagctcaaccgctc	1740
QY	1741	GATATTTGAAGGGAACATTATTAGAAGTCATATGTTCCAGCCATGACACGCTTCTCTCTA	1800
Db	1741	gatatltgaagggaacatttatgaagtcataatgtcttcagccatgacagctctctctcta	1800
QY	1801	TTAAAAATCACGGTTGATGCTGATGTTGATACTACGTTGACATACGACGCTTATCCCT	1860
Db	1801	ttaaaaatcacggtgtgtatgtctgtatgtatctaaacglttgatcatcgaagccttatccct	1860
QY	1861	GTTCTCGTGGGATTCCTAATTTGAAATACGAGTTCTCAAGGACATGGAATGTTAATTGG	1920
Db	1861	gttctcgtctggagctccatctcaagatacgaatcttccaagagcaatvgaaatgttaattgg	1920
QY	1921	ACTACGATACAGCTATACAAATACAAAAAGAGCCACGCACTTGGACACAAAACAGATT	1980
Db	1921	actacgatacagctatacctaataatacaaaaagagcgcaagcgaacttggaccaaaacgattt	1980
QY	1981	GTTCCGACCCCCGAAGAAATCGGCTTAGTAGCAATACCTTAGGGAGCTTTACT	2040
Db	1981	gttcccgagccccgaagaataatctcgcttgatgtacataacctatggggagcttact	2040
QY	2041	GACATTCGCTCTTCGAACAGCTTGTAGAGATGGCGCAACTGGTATGGAACACAAACA	2100
Db	2041	gacatctgctcttcgcaacagcgtctgtagaagaatcggcgcaacggtatggaacaacaaca	2100
QY	2101	GGTTTCTGGGTTTCCCTCATGACGAACCTTCTGCAATAGACTGTGAGATGAATAATGCA	2160
Db	2101	ggtttctggggttctctcatgacgaaccttctgcataagaatcggagatgaaatctgcaaa	2160
QY	2161	GGCTTCGCTCATACCTTGAGAGGCTACGTCATCGGTGGAAGTCTCACACTCTTAAGAC	2220
Db	2161	ggcttcgctcatacctcttgagagctacgctcatcgtggaagtgctcaacctctaaagac	2220
QY	2221	GACCATTTTACCTTGGGTTCTGCAATCTTGGTGCATAGACAAAGATGTTTATCCCT	2280
Db	2221	gacctatttaccttgggtctgctgcacatcccttgcctagagacaagaatctgtttatcgct	2280
QY	2281	CACAACAACCTCTAGAACCTTAGGCTGGGAACCTTATTTCTTACACACTCTCATACCTTAA	2340
Db	2281	cacaacaacctctagaacctagcgltggaacttatcttctaagcactctcataacctataaa	2340
QY	2341	CCCCAACAATTTTGAGATTAGAGAGACGAAGTTTCTGAAATCAGCTATAGAAAAATTC	2400
Db	2341	ccccaaacctatttgatagattagaagaagcaaaagtttcttgatcagctatagaaaaaatlc	2400
QY	2401	CCTTAGGGAATTCCTCCATAGGCTTGGATGTGCCAAGTTTGCTTAGCACTTCCAGACAACGT	2460
Db	2401	cccttagggaaattccctcatagaccttggatgtgtccaagtttcttagcattcagacaacgct	2460
QY	2461	ATGGAACGCACTTACCTCATTTGCCAGATCCGAAGTCTTGGAGCAACGAGCTGTATA	2520
Db	2461	atggaacgcaactataactcatattgccaagaatccgaagttcttggagcaaaagatgtcata	2520
QY	2521	GCTGTGTGTATGGCGCTAAGCACTTCCTTTGTTTCCAAACCAACATCTCTTTTCAAG	2580
Db	2521	gctgtgtgtatggcgttagacaccttctcttgtcttcttccaaccacacctctcttcaag	2580
QY	2581	ACCTTCATTTCCACGATGAAAGTCCGAATGCTTTATGTATCAACAAATAGCTTCTTGAA	2640
Db	2581	accttctatccacagatgaaagtcgaaatcgtttatgtatatacacaataatagcttcttgaa	2640
QY	2641	AGCTCTAGTGTAGGCGCTGTGTTTACTATTGGAAAGGCTCTTAACCTCTGATTTCTGCTG	2700
Db	2641	agctctatgtaatgtagcggtgtttagtatactgaaagcgcttcaaccttgcattcccggt	2700
QY	2701	GGTCCGAAATTTCTGCAAGGCGGATATCGAGATTTCTTCAACCTATGATCTCTCAGGATT	2760

[illegible]

RESULT 2  
 AAC81914/c  
 ID AAC81914 standard; DNA; 273254 BP.  
 XX  
 AC AAC81914;  
 XX  
 DT 27-FEB-2001 (first entry)  
 XX  
 DE Chlamydia pneumoniae genome DNA.  
 XX  
 KW Genome; diagnosis; vaccine; ds.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO200027994-A2.  
 XX  
 PD 18-MAY-2000.  
 XX  
 PF 12-NOV-1999; 99WO-US26923.  
 XX  
 PR 12-NOV-1998; 98US-0108279.  
 PR 08-APR-1999; 99US-0128606.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Stephens R, Mitchell W, Kalman S, Davis R;  
 XX  
 DR WPI: 2000-376516/32.  
 XX  
 PT Isolated nucleic acid for use in diagnostic and analytical methods  
 CC encodes genomic sequence of Chlamydia pneumoniae -  
 CC  
 PS Claim 2; Page 128-320; 320pp; English.  
 XX  
 CC This invention describes a novel nucleic acid (N1) encoding a Chlamydia  
 CC pneumoniae protein (P1), given in the specification. The isolated nucleic  
 CC acid is useful for diagnostic and analytical methods, such as,



DR P-PSDB; AAW88417.

PT Species-specific test for identifying mammals infected with  
PT *Chlamydia pneumoniae* - comprises detecting antibodies specific for  
PT outer membrane proteins of *C. pneumoniae* or nucleic acids encoding  
PT these proteins

PS Claim 6; Page 35-40; 115pp; English.

This DN sequence codes for the novel 98.9kDa surface exposed protein Ompd (see AAM68417) of the human respiratory pathogen *Chlamydia pneumoniae*. By generating antibodies against *C. pneumoniae* outer membrane complex, a polyclonal antibody (PAB 150) was obtained which reacted with outer membrane proteins. The antibody was used to identify the genes (see AA006816-77) encoding Ompd-Omp15 proteins (see AAM68417-88) in an expression library of *C. pneumoniae* DNA. The genes are situated in 2 gene clusters: Omp12.11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in the other, and encode polypeptides of about 89.6-100.3 kDa and about 56.1 kDa. The invention provides a new species specific test for identifying mammals (including humans) infected with *Chlamydia pneumoniae*. The test comprises detecting antibodies specific for Ompd-Omp15 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used in the diagnosis of *C. pneumoniae* infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting *in vivo* expression of antigens. The vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with *C. pneumoniae*.

**SQ** Sequence 3200 BP; 946 A; 708 C; 623 G; 923 T; 0 other;

**SQ** Sequence 3200 BP; 946 A; 708 C; 623 G; 923 T; 0 other;

Query Match	100.0%	Score 3200;	DB 20;	Length 3200;
Best Local Similarity	100.0%	Pred. NO. 0;		
Matches 3200; Conservative	0;	Mismatches	0;	Gaps 0;

Qy	1	CATGTGGAGAGGACGCTAAACGAGAAATTCGGATTTCATATAACCCATTATATATTA	60
Db	1	caatgltcgaagagagcactaacccaggaaattcgcatlctataaaccacttattatla	60
Qy	61	AATTCTTACTTGGCTCATATATAAATGAAATACACAGAGACTCAAGATATAAATTTCTGAC	120
Db	61	aattcttacttgcgtcatataaataagaacaccgaagagtgcaagataaanaattcttcgac	120
Qy	121	AGCGTTTGTGCATCTTAACTTGATTACTTATTTTGTTCATATATGATGGGAATAGT	180
Db	121	agcggtlcttgcatctttaaacttgatttacttatttggltctataltgaltggaatagt	180
Qy	181	TCCTTAATAAACAANAGCATTACCATGAACATCTGCATCTCTGGGTTTATGTTCTCTCC	240
Db	181	tccttaaaaaaacaagaagcattacaatgaagacttcgattcttcgggttttagttctctc	240
Qy	241	GTGTACTCTTTCATCTATCTACACTACAGTACTAGCTATACGAGAACTTTTATACCTGAT	300
Db	241	gtgttaagcttcttcataatgctacacagctacactagctaaacgaggaactttatatacttgat	300
Qy	301	GATAGCTTATATGGAATATATGATATTCAGTAACCTTACTCCAAAACTTCAGCACAAACA	360
Db	301	gataagctttaatggaatatatcgattcaagagcgttttaactccaaaacttcagccaataca	360
Qy	361	TATTTCTTAACAGAGAGATGTCTTCTTTTACAGAGCTTGAAAAAGCACTCCCTTATCTGAC	420
Db	361	tattcttcaacaggaagatgctcttcttctttaaagaccttggaaaagcactcccttatctgac	420
Qy	421	AGTTGTTTAAAGCAACACAGAGCAATCTTACCTCTTCTTGGAAGCGTCTATACCTTAACG	480
Db	421	agttgltttaagcaaacacaagacgaacacttaactcttcttgggaagcggtcatagctttaaag	480
Qy	481	TTTGGCTTTATAGATGCTGCACTCATGACAGGTGCTGCTGCATCTTACACAGCAAAATAG	540
Db	481	tttggctttaagatgctggcacaatcagaagcagtgctctcatctcacaacagaataaag	540

QY	541	AACTTACCTTCACAGGGTTTTCCTTACAGTAGTTTGATTCCCTCCCTGACCAACGGTT	600
Db	541	aacttacccttcacagggtttcccttaacgtagtttgattccctccctcagcaacggtt	600
QY	601	ACTACAGCTCAGGGAACGCTTCTCCAGCAGGAGGGTAAATTAGAAATATTGCTAA	660
Db	601	actacagctcagggaacgcttctccagcagggaggttaatttagaaaatttcgttaa	660
QY	661	CTTGTAGTTCGTGGAAATTTTTCTACTGCAGATGTGGAGCTATCAAGAGCGTCTTC	720
Db	661	cttgtagttctggaattttcttctacgcagatggtggagctatacaaggagcgtcttc	720
QY	721	CTTTTAACTGGCACTTCGCGAGATGCTTTTAAAGTAACACTCTTCATCAACAAAGGA	780
Db	721	cttttaactggcacttcggaatgagatgacctttttaagtaacaactcttcaacaaggga	780
QY	781	GGAGCAATTCGTACTACAGCAGCGCTCGCATAGCAATTAACACAGTTATTGTAGATTG	840
Db	781	ggagcaatttgtactacagcagcgctcgcatagcaattaacacaggttatgtagattc	840
QY	841	CTATCTAATCATAAGCGTCTACGTACGAGGGCGCATGTGATGATGAAGCAACGTGATCTA	900
Db	841	ctatctaatacataagcgtctacgtacgagggcgcatatgatagaagaagcagctcgatac	900
QY	901	TCGAACAACAATTTTCTATATTTTGAAGGGAATGACGCAAACTCTGCGGTGCGATC	960
Db	901	tcgaaacaacaattctatatatttgaagggaatgacgcaaaactctgsggtgscatc	960
QY	961	TGCAACACCAAGCGGAGTGCATCTCTGAACTGATATCTTATACATATAGACTGTGATC	1020
Db	961	tgcacaaccaagcgagtgagatctctctgaactgataatcttcaacaataagactctgac	1020
QY	1021	TTTGCTTCAACGTAGCAGGAAGCAAGGGGGGCGCATCCATCTTAAAGGTAGCCCTT	1080
Db	1021	tttgcttcaaacgtagcaggaagcaaggcggtggcgccatccaigtctaaagaactagccctt	1080
QY	1081	TCCTCTGGAGGCTTTACAGAGTTTCTACGAATATGTCTCATCAGCAACTCCTAAGGGG	1140
Db	1081	tcctctggagcgctttacagagtttctacgaataatgctctcatcagcaactcctaagg	1140
QY	1141	GSTGCTATCACATCGATGAGTCTCAGGAGCTGATGCTTTGCAAGACAGAGAAACATT	1200
Db	1141	gstgctatcacatcgatgagtcctcagagagctcagtccttcctgcagaagcaggaacaatt	1200
QY	1201	ACCTTTGTAAAGAAATPACCTTTACCAACCAACGGAAGTACCGATCTCTGTAAGCTATGCG	1260
Db	1201	acctttgtaaagaaatpacccttttaccaaccaacggaagtaccgatactactctaacgtatg	1260
QY	1261	ATCAACATAGGAAGTAAACGGGAATTACGGGAATTACGGGCTGCTAAAAATCATACATT	1320
Db	1261	atcaacaataggaagttaaaccgggaattacgggaattacggctgctaaaaatcatacaatt	1320
QY	1321	TTCTCTCTTGTGATCCCATTCACGCTTCGAAGGAACGTCATCAGACGTATTGGAAGTAAATAC	1380
Db	1321	ttctctctatgtgccatccaactctcgaaggaactctcagcaggtatgtagaataataaac	1380
QY	1381	GGCTCTGGGGAGGCTCTCAATCCATATCAAGAGCAATTCATTATTCTTGAGAAAACCTA	1440
Db	1381	ggctctggggaggtctcaatccaataccaagaagaagatctatcttcttgagaaaacctta	1440
QY	1441	ACAGACAGTGAACCTTAAGTTGGCGACAAATTAAATCTTCATTCACGACGACGATCC	1500
Db	1441	acagacagtgaaccttaaaagtgcgacaatttaaaactcttcatccaagcagcgtctcc	1500
QY	1501	CTATCCGAGAGGAAGTTATTGCTACAAAAGGAGTACATTTAGAGAGCAACGAGCTTCTCT	1560
Db	1501	ctatccgaggaaggttatgtctatacaaaaggagtgatcaactttagagcgacgagcttctc	1560
QY	1561	CAGAAGCGCGGTCTCTCTCTCGGCATGTGATTCAAGGAACGACATTATCAACTACAGCTGGG	1620
Db	1561	caagaagcgcggttctctctctccgcatagtattcaggaagacattatacaactacagctgg	1620



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2001, 13:17:33 ; Search time 800.41 Seconds  
(without alignments)  
2510.320 Million cell updates/sec

Title: US-09-446-677B-1  
Perfect score: 3200  
Sequence: 1 CAATGTCAGAGACACTA.....TTCTATTGACTTCTTTT 3200

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues  
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDSL/gcgdata/geneseq/NA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/NA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/NA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/NA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/NA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/NA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/NA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/NA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/NA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/NA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/NA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/NA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/NA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/NA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/NA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/NA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/NA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/NA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/NA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/NA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/NA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3200	100.0	3200	20	AA06816	Chlamydia pneumoni
2	3200	100.0	273254	21	AA081914	Chlamydia pneumoni
3	3199	100.0	1230025	20	AA091990	Nucleotide sequenc
4	2950	92.2	2950	21	AAA30847	Chlamydia antigen
5	2748	85.9	2784	21	AAA30848	Chlamydia antigen
6	555	17.3	3000	21	AAA30853	Chlamydia antigen
7	551.2	17.2	2793	20	AA06823	Chlamydia pneumoni
8	525.4	16.4	2790	21	AAA30854	Chlamydia antigen
9	508	15.9	2957	21	AA061509	DNA encoding the C
10	498.4	15.6	2757	20	AA06821	Chlamydia pneumoni
11	465.8	14.6	2787	20	AA06820	Chlamydia pneumoni

12	465.8	14.6	3000	21	AA027021	Chlamydia pneumoni
13	458.6	14.3	273254	21	AA081914	Chlamydia pneumoni
14	456	14.2	2815	20	AA06817	Chlamydia pneumoni
15	441.6	13.8	2950	21	AAA30849	Chlamydia antigen
16	433.6	13.6	3050	21	AAA48838	Chlamydia pneumoni
17	432.4	13.5	3000	20	AA06828	Chlamydia pneumoni
18	425.2	13.3	2950	21	AAA30851	Chlamydia antigen
19	423.2	13.2	2787	21	AA06822	Chlamydia pneumoni
20	422.8	13.2	2784	21	AAA30852	Chlamydia antigen
21	405.2	12.7	2808	21	AAA48839	Chlamydia pneumoni
22	382	11.9	2784	21	AAA30850	Chlamydia antigen
23	341	10.7	2526	20	AA06819	Chlamydia pneumoni
24	341	10.7	2526	21	AA06819	Chlamydia pneumoni
25	341	10.7	2750	21	AA06850	C. pneumoniae CPN1
26	323.2	10.1	3050	21	AA061508	DNA encoding the C
27	307.6	9.6	2838	20	AA06827	Chlamydia pneumoni
28	300	9.4	3000	20	AA06828	Chlamydia pneumoni
29	255.8	8.0	2950	21	AAA50035	DNA encoding Chlam
30	252.6	7.9	3052	20	AA06818	Chlamydia pneumoni
31	191.8	6.0	840	20	AA06824	Chlamydia pneumoni
32	184.6	5.8	1830	20	AA06830	Chlamydia pneumoni
33	143.6	4.5	3150	21	AA027342	Chlamydia POMP18
34	142	4.4	2950	21	AAA30849	Chlamydia antigen
35	113.2	3.5	936	22	AA058254	Oligonucleotide D1
36	113.2	3.5	936	22	AA058254	Oligonucleotide D1
37	113.2	3.5	936	22	AA058257	Oligonucleotide D2
38	113.2	3.5	936	22	AA058262	Oligonucleotide D2
39	113.2	3.5	936	22	AA058262	Oligonucleotide D1
40	113.2	3.5	936	22	AA058255	Oligonucleotide D1
41	112	3.5	936	22	AA058252	Oligonucleotide D1
42	112	3.5	936	22	AA058254	Oligonucleotide D1
43	112	3.5	936	22	AA058257	Oligonucleotide D1
44	112	3.5	936	22	AA058259	Oligonucleotide D2
45	112	3.5	936	22	AA058262	Oligonucleotide D2

## ALIGNMENTS

RESULT 1	
AA06816	AA06816 standard; DNA: 3200 BP.
XX	
AC	AA06816:
XX	
DT	26-APR-1999 (first entry)
XX	
DE	Chlamydia pneumoniae surface exposed protein Omp4 DNA.
XX	
KW	Omp4: outer membrane protein 4; surface exposed protein; antigen;
KM	infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX	
OS	Chlamydia pneumoniae.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
XX	205..2991
XX	/*tag= a
PN	WO9858953-A2.
PD	30-DEC-1998.
XX	
PF	19-JUN-1998; 98WO-DK00266.
XX	
PR	23-JUN-1997; 97DK-0000744.
XX	
PA	(BIRK/) BIRKELUND S.
XX	(CHR1/) CHRISTIANSEN G.
PI	Birkelund S, Christiansen G, Knudsen K, Madsen A,
XX	Mygind P.
XX	WPI: 1999-105610/09.

**THIS PAGE BLANK (USPTO)**

---





QY	2991	GATTCGCAAAACCCCAAGTCTCTTCAGGAGAGTTTCTCATCTTTTGGGAAATATTTG	3050
Db	370	GATTGGTAATTAAGTTTATTTTATTTTATATATCATTTTAAAACTTTCATATAAAGATTCC	429
QY	3051	CTATAGGAAATGCTTTCCTTCGCAACTGTAAAAATATACATTTTGTCCCTCTCAAAAAG	3110
Db	430	AGAAATAATTAATATTTTTTTTGTAAACTCGGAAATTCAGCAATTTTCTTTTAATCATATTC	489
QY	3111	ATTCTTTTAATAATTTCTGAGTATATAATTTTATTTTAAAAAGAGTAAATAATTAATGA	3170
Db	490	ATTATATCTCATATATACATAAATCTCTTTTATTTTAAATCATATAAAAAGCAATGAAGA	549
QY	3171	CAATATCATATCTTATTT 3188	
Db	550	TATTAATCTTCTCATACT 567	

OY 1888 TAGGATTCAGAGCAATGATGTTAATGGACTACGATACGATCAATCAATAA 1947  
DB 1492 AGGGGATCTATGCTCTTTGGATTTGGGACAGAAAGTCCAAATTTACCACTTATTCATGA 1551  
OY 1948 GAGGCCACGG 1957  
DB 1552 GTCACCTACTG 1561

RESULT 10  
PCT-US93-10520-1  
Sequence 1, Application PC/TUS9310520  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSER: CREATIVE BIOMOLECULES, INC.  
STREET: 45 SOUTH STREET  
CITY: HOPKINTON  
STATE: MA  
COUNTRY: USA  
ZIP: 01748  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10520  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/667,274  
FILING DATE: 11-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/752,764  
FILING DATE: 30-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/753,059  
FILING DATE: 30-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/752,857  
FILING DATE: 30-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/923,780  
FILING DATE: 31-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/922,813  
FILING DATE: 31-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER ESO, EDMUND R  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: CRP-076PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (508)435-9001  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1674 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 69..1268  
OTHER INFORMATION: /note= "mop3-pp"  
PCT-US93-10520-1

Query Match 1.2%; Score 39.6; DB 5; Length 1674;  
Best Local Similarity 50.5%; Pred. No. 0.45; Mismatches 96; Conservative 0; Indels 0; Gaps 0;

OY 1768 CATATGTCAGCCATGACAGCCTTCTCTATTAATAATCAGGTTGATGTT 1827  
DB 1372 CACATTCACAGCTTGCGCCCTCTGTTCTTTTGCAAGGCTGGAAGATGCTCTAGT 1431  
OY 1828 GATTAACCTTGACATCAGCAGCCTTATCCCTGTTCTGAGAGCTTAATTCAGAA 1887  
DB 1432 TATAACCCCTGAGCTCCTTTGGATTTGGGACAGAAAGTCCAAATTTACCACTTATTCATGA 1491  
OY 1888 TAGGATTCAGAGCAATGATGTTAATGGACTACGATACGATCAATCAATAA 1947  
DB 1492 AGGGGATCTATGCTCTTTGGATTTGGGACAGAAAGTCCAAATTTACCACTTATTCATGA 1551  
OY 1948 GAGGCCACGG 1957  
DB 1552 GTCACCTACTG 1561

RESULT 11  
US-07-991-867B-41  
Sequence 41, Application US/07991867B  
Patent No. 5476781  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/991,867B  
FILING DATE: 12-DEC-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO 92/14818  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,685  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UF114.C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1689 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-07-991-867B-41

Query Match 1.2%; Score 39.6; DB 1; Length 1689;  
Best Local Similarity 50.0%; Pred. No. 0.45; Mismatches 99; Conservative 0; Indels 0; Gaps 0;

[illegible]

```

RESULT      8
US-08-901-200A-1
: Sequence 1, Application US/08901200A
: Patent No. 5854071
:
GENERAL INFORMATION:
:
: APPLICANT: OPPERMAN, HERMANN
: APPLICANT: OZKAYNAK, ENGIN
: APPLICANT: KUBERASAMPAH, THANGAVEL
: APPLICANT: RUEGER, DAVID C.
: APPLICANT: PANG, ROY H.L.
: APPLICANT: COHEN, CHARLES M.
: TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
: STREET: 45 SOUTH STREET
: CITY: HOPKINTON
: STATE: MA
: COUNTRY: USA
: ZIP: 01748
:
COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/901,200A
: FILING DATE: 28-Jul-1997
:
CLASSIFICATION: 530
:
ATTORNEY/AGENT INFORMATION:
: NAME: MEYERS, THOMAS C.
: REGISTRATION NUMBER: 36,989
: REFERENCE/DOCKET NUMBER: CRP-076DV2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (508) 435-9001
: TELEFAX: (508) 435-6951
:
INFORMATION FOR SEQ ID NO: 1:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1674 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
MOLECULE TYPE:
:
FEATURE:
:
: NAME/KEY: CDS
: LOCATION: 69..1265
: OTHER INFORMATION: /note= "mop3-pp"
:
: US-08-901-200A-1

```

[illegible][illegible]

```

RESULT          9
US-09-219-391-1
Sequence 1, Application US/09219391
Patent No. 6153583
GENERAL INFORMATION:
APPLICANT: OPPERMANN, HERMANN
APPLICANT: OZKANYAK, ENGIN
APPLICANT: KUBERASAMPATH, THANGAVEL
APPLICANT: RUEGER, DAVID C.
APPLICANT: PANG, ROY H.L.
APPLICANT: COHEN, CHARLES M.
TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
STREET: 45 SOUTH STREET
CITY: HOPKINTON
STATE: MA
COUNTRY: USA
ZIP: 01748
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/219,391
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/901,200
FILING DATE: 28-Jul-1997
ATTORNEY/AGENT INFORMATION:
NAME: MEYERS, THOMAS C.
REGISTRATION NUMBER: 36,989
REFERENCE/DOCKET NUMBER: CRP-076DV2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 435-9001
TELEFAX: (508) 435-6951
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1674 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
FEATURE:
NAME/KEY: CDS
LOCATION: 69..1265
OTHER INFORMATION: /note="mob3-pp"
US-09-219-391-1

```

[illegible]





TELEFAX: (216)241 0816  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOHETICAL: NO  
US-09-031-897-1

Query Match 1.3%; Score 41.6; DB 3; Length 2483;  
Best local Similarity 59.5%; Pred. No. 0.16;  
Matches 88; Conservative 0; Mismatches 59; Indels 1; Gaps 1;

QY 3032 ACTTTAGGGAATATTGCTATAGGAATGCTTCCTGCAACGTATAAATAACAT 3091  
DB 1386 ATTTTCAGATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1327  
QY 3092 TTGTCCTCTTCAAAAAGATTCTTTATATTTCTACTTATATTTTAAAAA 3151  
DB 1326 TTCTACCTCTATTGACATATTTCCAGTATGAAATTC-ATTCGCAATTTTATTTCTAAA 1268  
QY 3152 CAGTTAAATTAATTAAGACATATCT 3179  
DB 1267 TATTATCTAATTTATCTAATAAATAAT 1240

RESULT 4  
US-08-731-722-4  
Sequence 4, Application US/08731722  
Patent No. 5961971  
GENERAL INFORMATION:  
APPLICANT: Martin, Frank N.  
TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens  
TITLE OF INVENTION: by Pythium oligandrum  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1.  
CITY: Gainesville  
STATE: FL  
COUNTRY: US  
ZIP: 32606-6669  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/731,722  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Whitlock, Ted W.  
REGISTRATION NUMBER: 36,965  
REFERENCE/DOCKET NUMBER: UF-161  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 352-375-8100  
TELEFAX: 352-372-5800  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: 17-1  
US-08-731-722-4

Query Match 1.3%; Score 40.4; DB 2; Length 1218;  
Best local Similarity 50.5%; Pred. No. 0.24;  
Matches 98; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 3007 TAGTCTCTAGGAGTTTTCATACCTTTAGGAAATATTCATAGGAATGCTTT 3066  
DB 547 TAAATATTGAAGACCTTTACTATATTTCTAATATATATCGAATTTAAACATGAGAC 606  
QY 3067 CCTTGCAAACTGTAAAAAATAACATTGTCCTCCCTCAAAAAAGATTCTTAAATAAT 3126  
DB 607 CCGTAGAACCTTTAATTAATGCTATAGAAATTTTCATAGAGAAAGCTTTTATTTT 666  
QY 3127 TCTAGTTATATTTTATTTTAAAAACGTTAATATTAATAGACATATCTATTCTTA 3186  
DB 667 TAATAGAAATTTATTTCTATTATTACTTTAATATAGAAATTAATATTAATTTATA 726  
QY 3187 TTGACTTCTTTT 3200  
DB 727 TAAATATATTCTT 740

RESULT 5  
US-08-323-170B-1  
Sequence 1, Application US/08323170B  
Patent No. 5733772  
GENERAL INFORMATION:  
APPLICANT: Williamson, Kim C.  
TITLE OF INVENTION: Cloning and Expression of Plasmodium  
TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pf230  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/323,170B  
FILING DATE: 13-OCT-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/010,409  
FILING DATE: 29-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Quine, Jonathan A.  
REGISTRATION NUMBER: P-41,261  
REFERENCE/DOCKET NUMBER: 015280-113100US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9636 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 149..9556  
US-08-323-170B-1

Query Match 1.2%; Score 40; DB 1; Length 9636;  
Best local Similarity 52.4%; Pred. No. 0.74;  
Matches 88; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

DB 1386 ATTTTCAGATATATATATACATATATATATAGATGATTAAGTAACTATATATATGAT 1327  
QY 3092 TTGTCCTCTTCAAAAAGATTTCTTTATATATCTAGTATATATTTATTTAAAAA 3151  
DB 1326 TTCTACCTCTTATTAACATATTTCCAGATATGAAATTC-ATTCTCAATTTATTTCTAAA 1268  
QY 3152 CAGTTAAATATTAATAGACATATATCT 3179  
DB 1267 TATTTATCTAATTTATCTAAAAAATAT 1240

RESULT 2  
US-08-946-617-1/c  
Sequence 5, Application US/08911320A  
Patent No. 5869633  
GENERAL INFORMATION:  
APPLICANT: INCYTE PHARMACEUTICALS, INC.  
TITLE OF INVENTION: THROMBIN RECEPTOR HOMOLOG  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/911.320A  
FILING DATE: August 14, 1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/467,125  
FILING DATE: 6-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0041-1 DIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-845-4170  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
Sequence 1, Application US/08946617  
Patent No. 5869634  
GENERAL INFORMATION:  
APPLICANT: Lambowitz, Alan M  
APPLICANT: Zimmerly, Steven  
APPLICANT: Guo, Huatao  
APPLICANT: Yang, Jian  
TITLE OF INVENTION: Nucleotide Integrase Preparation  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calfee, Halter & Griswold  
STREET: 800 Superior Avenue  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: USA  
ZIP: 44114  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/946,617  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldick, Mary E  
REGISTRATION NUMBER: 34,829  
REFERENCE/DOCKET NUMBER: 22727/00127  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (216) 622-8200  
TELEFAX: (216) 241-0616  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
US-08-946-617-1

Query Match 1.3%; Score 41.6; DB 2; Length 2483;  
Best Local Similarity 59.5%; Pred. No. 0.16;  
Matches 88; Conservative 0; Mismatches 59; Indels 1; Gaps 1;

QY 3032 ACTTTAGGGAATATTTGCTATAGGAATGCTTCTTCAAGCTGTAATAATACAT 3091  
DB 1386 ATTTTCAGATATATATATTTACATCTATATATTTAGATATTAAGTATATATGAT 1327  
QY 3092 TTGTCCTCTTCAAAAAGATTTCTTTATATATCTAGTATATATTTATTTAAAAA 3151  
DB 1326 TTCTACCTCTTATTAACATATTTCCAGATATGAAATTC-ATTCTCAATTTATTTCTAAA 1268  
QY 3152 CAGTTAAATATTAATAGACATATATCT 3179  
DB 1267 TATTTATCTAATTTATCTAAAAAATAT 1240

RESULT 3  
US-09-031-897-1/c  
Sequence 1, Application US/09031897  
Patent No. 6027895  
GENERAL INFORMATION:  
APPLICANT: Lambowitz, Alan  
APPLICANT: Mohr, Georg  
APPLICANT: Zimmerly, Steven  
APPLICANT: Guo, Huatao  
TITLE OF INVENTION: Methods Cleaving DNA with Nucleotide  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calfee, Halter & Griswold  
STREET: 800 Superior Avenue, Suite 1400  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: US  
ZIP: 44114  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/031,897  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Docherty, Pamela A.  
REGISTRATION NUMBER: 40,591  
REFERENCE/DOCKET NUMBER: 24671/00105  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (216) 622-8416

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2001, 13:30:07 ; Search time 268.88 Seconds  
(without alignments)  
2253.033 Million cell updates/sec

Title: US-09-446-677B-1  
Perfect score: 3200  
Sequence: 1 CAATGTCGAAGAGACACTA.....TTCTATTGACTTCTTTT 3200

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/2/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/lna/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/lna/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/lna/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	41.6	1.3	2483	1	US-08-526-964-1 Sequence 1, App11
C 2	41.6	1.3	2483	2	US-08-946-617-1 Sequence 5, App11
C 3	41.6	1.3	2483	3	US-09-031-897-1 Sequence 1, App11
C 4	40.4	1.3	1218	2	US-08-731-722-4 Sequence 4, App11
C 5	40	1.2	9636	1	US-08-323-170B-1 Sequence 1, App11
C 6	39.6	1.2	1674	1	US-08-480-528B-1 Sequence 1, App11
C 7	39.6	1.2	1674	1	US-08-479-666-1 Sequence 1, App11
C 8	39.6	1.2	1674	2	US-08-901-200A-1 Sequence 1, App11
C 9	39.6	1.2	1674	3	US-09-219-391-1 Sequence 1, App11
C 10	39.6	1.2	1674	5	PCT-US93-10520-1 Sequence 1, App11
C 11	39.6	1.2	1689	2	US-07-991-867B-41 Sequence 41, App1
C 12	39.6	1.2	1689	2	US-08-544-332-41 Sequence 41, App1
C 13	39.6	1.2	8457	1	US-07-991-867B-1 Sequence 1, App11
C 14	39.6	1.2	8457	2	US-08-544-332-1 Sequence 1, App11
C 15	39.2	1.2	2960	3	US-08-913-842-3 Sequence 3, App11
C 16	38.4	1.2	1415	1	US-08-413-118-126 Sequence 126, App
C 17	38.4	1.2	1415	3	US-08-473-446-126 Sequence 14, App1
C 18	38.4	1.2	7218	1	US-08-232-463-14 Sequence 14, App1
C 19	37.6	1.2	837	4	US-08-998-416-288 Sequence 28, App
C 20	37.4	1.2	7244	4	US-08-378-313-26 Sequence 26, App1
C 21	37.2	1.2	1241	1	US-07-593-657-6 Sequence 6, App11
C 22	37.2	1.2	1241	4	US-08-942-012B-3 Sequence 3, App11
C 23	37	1.2	662	4	US-08-998-416-185 Sequence 185, App
C 24	37	1.2	663	4	US-08-998-416-191 Sequence 191, App
C 25	37	1.2	665	4	US-08-998-416-937 Sequence 937, App
C 26	37	1.2	701	4	US-08-998-416-701 Sequence 701, App
C 27	37	1.2	711	4	US-08-998-416-786 Sequence 786, App

C 28	37	1.2	724	4	US-08-998-416-683 Sequence 683, App
C 29	37	1.2	732	4	US-08-998-416-1036 Sequence 1036, App
C 30	37	1.2	767	4	US-08-998-416-472 Sequence 472, App
C 31	37	1.2	782	4	US-08-998-416-224 Sequence 224, App
C 32	37	1.2	827	4	US-08-998-416-535 Sequence 535, App
C 33	37	1.2	828	4	US-08-998-416-538 Sequence 538, App
C 34	37	1.2	834	4	US-08-998-416-305 Sequence 305, App
C 35	36.8	1.1	1850	3	US-08-617-860B-32 Sequence 32, App1
C 36	36.8	1.1	4098	2	US-08-605-106-4 Sequence 4, App11
C 37	36.6	1.1	998	4	US-09-122-400B-5 Sequence 5, App11
C 38	36.6	1.1	2656	4	US-09-117-257-7 Sequence 7, App11
C 39	36.6	1.1	8700	2	US-08-392-625-16 Sequence 16, App1
C 40	36.6	1.1	8700	2	US-08-466-961A-16 Sequence 16, App1
C 41	36.6	1.1	8700	2	US-08-645-193B-18 Sequence 18, App1
C 42	36.4	1.1	5852	1	US-07-867-106-2 Sequence 2, App11
C 43	36.2	1.1	934	2	US-08-731-722-6 Sequence 6, App11
C 44	36.2	1.1	3095	6	Patent No. 5231168
C 45	36	1.1	2251	4	US-08-991-677-11 Sequence 11, App1

#### ALIGNMENTS

```
RESULT 1
US-08-526-964-1/c
; Sequence 1, Application US/08526964
; Patent No. 5698421
; GENERAL INFORMATION:
; APPLICANT: Lamberly, Alan M
; APPLICANT: Zimmerman, Steven
; APPLICANT: Guo, Huatao
; APPLICANT: Yang, Jian
; TITLE OF INVENTION: Nucleotide Integrase Preparation
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter & Griswold
; STREET: 800 Superior Avenue
; City: Cleveland
; STATE: Ohio
; COUNTRY: USA
; ZIP: 44114
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,964
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goitick, Mary E
; REGISTRATION NUMBER: 34,829
; TELEPHONE/DOCKET NUMBER: 22727/00127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8200
; TELEFAX: (216) 241-0816
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2483 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; US-08-526-964-1

Query Match 1.3%; Score 41.6; DB 1; Length 2483;
Best Local Similarity 59.5%; Pred. No. 0.16;
Matches 88; Conservative 0; Mismatches 59; Indels 1; Gaps 1;
QY 3032 ACCTTAGGAATATTGCTATAGGAATGCTTCTCGCAACGTGTAATAAATAACAT 3091
```

Best Local Similarity 42.18; Pred. No. 0.36;  
Matches 75; Conservative 29; Mismatches 74; Indels 0; Gaps 0;

[illegible]

Search completed: October 2, 2001, 03:06:06  
Job time: 51624 sec

```

/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOSI; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      380 a      100 c      132 g      281 t
ORIGIN

Query Match      1.5%: Score 48.2; DB 245; Length 893;
Best Local Similarity 58.2%: Pred. No. 0.3;
Matches 103; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

OY  2996 CTAACACCTCCCTAGCTTCTAGGAGTTTCATACCTTTAGGAAATATTGCTATA 3055
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  598 CTAACACTCTTATCTCTCTATATATGTGAGAAATTTGACACTTCTACGCTTTA 539

OY  3056 GGAATGCTTTCCTGCAAACTGTAATAATACATTGTCCTTCMAAAAGATTTC 3115
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  538 GTCATGCTTTTTTTTGTGATTTGTCAGTATATCATCTTCAT-TTTAATATATTTA 480

OY  3116 TTTTAATATTTCTAGTATATTTTATTTAAACAGTTAATATATATAGACA 3172
      ||||| ||||| ||||| ||| ||||| ||||| ||| ||||| ||||| |||||
DB  479 TTTTATTAATTTTATGTTTATTTTAAATTTTAAACAAATTAATTAATTAATA 423

RESULT 14
CNS017VX/c      964 bp      DNA      GSS      26-JUL-1999
LOCUS            Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION      BACN37015 of DrosBAC library from Drosophila melanogaster (fruit
                  fly), genomic survey sequence.
ACCESSION      AL108567
VERSION        AL108567.1 GI:5628871
KEYWORDS       GSS.
SOURCE         fruit fly.
ORGANISM       Plasmid Drosophila melanogaster
                Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE       1 (bases 1 to 964)
                Genoscope.
                Direct Submission
                Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
                BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
                - Web : www.genoscope.cns.fr)
                Determination of this BAC-end sequence was carried out as part of a
                collaboration with the European Drosophila Genome Project (EDGP) -
                http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
                library (Dros BAC) was made by Alain Billand at CEPH (Centre
                d'Etude du Polymorphisme Humain) with funding provided by a MC
                Project grant. The DNA was prepared from embryos by Alain Bucheton
                and Genevieve Payan. It has been constructed in the vector
                pBelobAC11.
FEATURES
source          Location/Qualifiers
                1..964
                /organism="Drosophila melanogaster"
                /plasmid="pBelobAC11"
                /db_xref="taxon:7227"
                /clone_lib="DrosBAC"
                /clone="BACN37015"

```

```

/note="end : SP6"
BASE COUNT      311 a      131 c      117 g      282 t      123 others
ORIGIN

Query Match      1.5%: Score 48.2; DB 219; Length 964;
Best Local Similarity 41.9%: Pred. No. 0.31;
Matches 80; Conservative 29; Mismatches 82; Indels 0; Gaps 0;

OY  3010 TTCCTCTAGGAGTTTCTCATCTTTAGGAAATATTTCTATAGGAATGCTTCT 3069
      || | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  925 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 866

OY  3070 TGCAACGTAAATAATACATTTGCTCCCTGCAAAAAGATTCTTTAATATTTCT 3129
      : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  865 WAAAAAATAATATATATATATATATATATATATATATATATATATATATAT 806

OY  3130 AGTATATAATTTTATTTTAAACAGTTAATTAATTAATAGCAATATCTATTG 3189
      : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  805 WATTAATTTTATACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 746

OY  3190 ACTTCTTTTCT 3200
      | : | | | | |
DB  745 AATATTTTCTY 735

RESULT 15
CNS0021J      1101 bp      DNA      GSS      03-JUN-1999
LOCUS            Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION      BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit
                  fly), genomic survey sequence.
ACCESSION      A1061936
VERSION        A1061936.1 GI:4940214
KEYWORDS       GSS.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
                Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE       1 (bases 1 to 1101)
                Genoscope.
                Direct Submission
                Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
                BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
                - Web : www.genoscope.cns.fr)
                Determination of this BAC-end sequence was carried out as part of a
                collaboration with the Berkeley Drosophila Genome Project (BDGP).
                The BDGP is constructing a physical map of the Drosophila
                melanogaster genome using these BACs. For further information
                please see http://www.fruitfly.org The BDGP Drosophila
                melanogaster BAC library was prepared by Kazutoyo Oosagawa and
                Aaron Mammosser in Pieter de Jong's Laboratory in the Department of
                Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
                NY. The library is named RPCI-98 and was constructed by partial
                EcoRI digestion of Drosophila DNA provided by the BDGP from the
                isogenic strain y2; cn bw sp. the same strain used for the BDGP's
                pl and EST libraries. A more detailed description of the library
                and how to order individual BAC clones, the entire library, or
                filters for hybridization from the BACPAC Resource Center can be
                found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source          Location/Qualifiers
                1..1101
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone_lib="RPCI-98"
                /clone="BACR05N11"
                /note="end : TET3"
BASE COUNT      631 a      7 c      28 g      289 t      146 others
ORIGIN

Query Match      1.5%: Score 48; DB 219; Length 1101;

```

LOCUS	CH016Y1	999 bp	DNA	GSS	26-Jul-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BACN17C05 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL107364				
VERSION	AL107364.1 GI:5627003				
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Plasmid Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 999)				
REFERENCE	Genoscope.				
AUTHORS	Direct Submission				
TITLE	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage				
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector peloBAC11.				
FEATURES	location/Qualifiers				
SOURCE	1..999 /organism="Drosophila melanogaster" /plasmid="peloBAC11" /db_xref="taxon:7227" /clone_1lb="DrosBAC" /clone="BACN17C05" /note="end : T7"				
BASE COUNT	253 a 113 c 94 g 144 t 395 others				
ORIGIN					
Query Match	1.5%; Score 48.6; DB 219; Length 999;				
Best Local Similarity	36.0%; Pred. No. 0.25;				
Matches	91; Conservative 30; Mismatches 132; Indels 0; Gaps 0;				
Y 2948	CAAGGAAGTACAAGTAGATGTGGTACCACAAACCCGATTCAGATGCTAAACTCCT 3007				
B 609	CACNANANNAAAAKAKTAKAKAAAAANANANAAACNNANAAATCTANNNNNNNAACNNCA 668				
Y 3008	AGTCTCTAGGAGATTTTCATCTTCTTGGGAAATATTTGCTATAGGGAATGCTTC 3067				
B 669	NNNANNNNAACCAATKTKNNNNKTNNNANKANNANAAATATACGYTTTAAWTTTMAA 728				
Y 3068	CTTGCAACCTGTAAAAAATTAACATTTTGCCCTCTTCAAAAAAGATTTCTTTAATAATTT 3127				
B 729	ATTTTAATTTTTTYYBDRKACMCCCAAAACCCGCCCAAAAAABAAATTAATAATTTT 788				
Y 3128	CTAGTTAAATTTTATTTTAAACAGTTAATAATATATATAGACAATATATCTATCTTAT 3187				
B 789	CTTAATATATATATTTTAAAHAYTAATTTTNNHCAAAAAAMATTAATAATATTT 848				
Y 3188	TGACTTCCTTTT 3200				
B 849	TYYTTTTTYYAT 861				
RESULT 12					
LOCUS	C23700				
DEFINITION	C23700 759 bp mRNA EST 16-OCT-1997				
ACCESSION	C23700				
VERSION	C23700.1 GI:2073212				
KEYWORDS	EST.				
SOURCE	Dictyostelium discoideum.				
ORGANISM	Dictyostelium discoideum				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 759)			
	Ikayota; Mycetozoa; Dictyostellidae; Dictyostellium.			
	Shimizu, H. and Urushihara, H.			
	Sexual-cDNA in CSM			
	Unpublished (1997)			
	Contact: Hideo Urushihara			
	Institute of Biological Sciences			
	University of Tsukuba			
	3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan			
	Email: d402huesakura.cc.tsukuba.ac.jp.			
	Location/Qualifiers			
	1..759			
	/organism="Dictyostellium discoideum"			
	/strain="KAX3"			
	/db_xref="taxon:44699"			
	/clone="FC-AB19"			
	/clone.lib="Dictyostellium discoideum FC (H.Urushihara)"			
	ORIGIN			
	BASE COUNT	339 a	110 c	79 g 231 t
	Query Match	1.5%;	Score 48.2;	DB 156; Length 759;
	Best Local Similarity	57.3%;	Pred. No. 0.29;	Indels 1; Gaps 1
	Matches 106; Conservative	0;	Mismatches 78;	
QY	10	AGAGAGCATTACAGGAAATATGCGATTCATTAACCCACCTTATTAATTCCTAC	69	
Db	572	AAAAAGATTAAGGAAAGTTGAATATATTAATAATTCATATTAATTTT	513	
QY	70	TTCGCGCATATAAATATGAAACTGAGAGTCAGATTAATTCCTGACAGCTTT	129	
Db	512	TTTTATTAATAAAGGAAAGGAAAGGAAAGGAAAGATTAATGTT-TAATTTGTTTAT	454	
QY	130	GTCATCTTAACTTGAATTTACTATTTTCTTCTATATGATGCGAATGTCCTATAA	189	
Db	453	CTTTTCTTTTCTTTGATTTCTTCTTTTCTTTTCTTTGTCATCTTTCTTTTGA	394	
QY	190	AACAA 194		
Db	393	AACCA 389		
RESULT 13	AZ535391/c	893 bp	DNA	03-NOV-2000
LOCUS	ENTB0887R	Entamoeba histolytica	Sheared DNA	Entamoeba histolytica
DEFINITION	genomic, DNA sequence.			
ACCESSION	AZ535391			
VERSION	AZ535391.1	GI:11092338		
KEYWORDS	GSS.			
SOURCE	Entamoeba histolytica.			
ORGANISM	Entamoeba histolytica			
REFERENCE	1 (bases 1 to 893)			
AUTHORS	Loftus, B., Van Aken, S. and Fraser, C.			
TITLE	Determination of clone end sequences from Entamoeba histolytica			
JOURNAL	HML:IMSS sheared DNA library			
COMMENT	Unpublished (2000)			
	Contact: Brendan J Loftus			
	Department of Eukaryotic Genomics			
	The Institute for Genomic Research			
	9712 Medical Center Dr., Rockville, MD 20850, USA			
	Tel: 301 838 0200			
	Fax: 301 838 0208			
	Email: b.loftus@tigr.org			
	Clones are derived from the Entamoeba histolytica HML:IMSS sheared			
	DNA library			
	Seq primer: M13-Reverse			
	Class: Shotgun			
	High quality sequence start: 674			
	High quality sequence stop: 811.			
	Location/Qualifiers			
	1..893			



Matches	74: Conservative	27: Mismatches	66: Indels	0: Gaps
Oy	3034	TTTTGAGCAATATTTCGATATGAGCAATGCTTCCTCGCAACGTAAATAATCAATT	3093	
Db	1165	TTAKATAAATAAAATWATTTTMMWRATWATWTAAMTMAAATAAAAAAAATTTMMW	1106	
Oy	3094	GTCCTCTTCACAAAAGATTTCTTTTAATTAATTTTCAGTATATATTTTATTTAAACA	3153	
Db	1105	TWMAATTTTAAATWATTTTTTTTTTTTTTTTTAAAMAAATATATATTTATTTAATTA	1046	
Oy	3154	GTTAAATATTAATATAGACATATCTATTTCTATTGACTTCCTTTT	3200	
Db	1045	TTTAATTTATTTAATTTTAAATTTTMMWATATATMAAATTTTTTT	999	
RESULT	9			
CNS00DKY				
LOCUS				
DEFINITION	CNS00DKY	928 bp	DNA	GSS
ACCESSION				04-JUN-1999
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
FEATURES				
source				
BASE COUNT				
ORIGIN				
Query Match	1.5%	Score 49.2;	DB 219;	Length 928;
Best Local Similarity	49.2%;	Pred. No. 0.117;		
Matches	60;	Conservative 22;	Mismatches 40;	Indels 0;
			Gaps	0;
Oy	3079	TAAAAAATACATTTGTCCTCTCAAAAAAGATTTCTTTAATATATTTCTAGATTAAT	3138	
Db	521	WAAAAAATAATTTTTTTTAAATAAATAATTTTTTTTTTTTAAAT	580	
Oy	3139	TTTATTTAAAAAGCTTAATAATTAATADAGACATATCTCTCTATTCGCTCTTT	3198	
Db	561	TWTATTTAAATTTAAATWATTTTATTAATWTAATAATATATTTATTAATATTTWT	640	

QY	3199	TP	3200
DB	641	TW	642
RESULT	10		
LOCUS	CNS036CC/C		
DEFINITION	CNS036CC	637 bp	DNA
ACCESSION	AL229845		
VERSION	AL229845.1		
KEYWORDS	GSS: genome survey sequence.		
SOURCE	Tetradodon nigroviridis genome survey sequence.		
ORGANISM	Tetradodon nigroviridis		
REFERENCE	GSS: genome survey sequence.		
AUTHORS	Tetradodon nigroviridis		
TITLE	Characterization and repeat analysis of the freshwater pufferfish Tetradodon nigroviridis		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 637)		
AUTHORS	Roeck-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizeses,C., Mincker,P., Brotlier,P., Quetier.F., Saurin,W. and Weissenbach,J.		
TITLE	Human gene number estimate provided by genome wide analysis using Tetradodon nigroviridis DNA sequence		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 637)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases		
COMMENT	This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetradodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/Tetradodon">http://www.genoscope.cns.fr/Tetradodon</a> .		
FEATURES	Location/Qualifiers		
Source	1..637		
	/organism="Tetradodon nigroviridis"		
	/db_xref="taxon:99883"		
	/clone="215L21"		
	/clone_1lb="G"		
	/note="Genoscope sequence ID : C0AG215CF11SP1-end : PUC-Orl"		
BASE COUNT	180 a	95 c	80 g
ORIGIN	224 t	58 others	
Query Match	1.5%	Score 48.6:	DB 220; Length 637;
Best Local Similarity	47.2%	Pred. No. 0.22;	
Matches	75; Conservative	20; Mismatches	64; Indels 0; Gaps 0;
QY	3042	AAAAATTGCTATAGGAGATGCTTCCTGCACAACTGTAATAAATAATACATTGTCCCTCT	3101
DB	558	AAAAATTTTWTWMAAAAAATTTATATTTTTTTTTTMAAAAAAAATTTTTTTTAA	499
QY	3102	TCAAAAAGATTCTTTTAAATATTTCTAGTTAATTTTAAATTTTAAACAGTTAAATA	3161
DB	498	AAAAAAAATTTTTTTTMAAMWTTTTTTTWTWTTTTTTTTTTTMAAAAAAAATTTTTT	439
QY	3162	ATTAAATAGCAATATCATCTTATTCAGTTCTTTTT	3200
DB	438	TTTTTTTATWMAAAATKGGCGAAATTTTTTT	400
RESULT	11		
CNS016Y1			





VERSION AL069440.1 GI:4949583  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1101)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
COMMENT - Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila genome project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
FEATURES  
Source  
1..1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR29P01"  
/note="end : TET3"  
BASE COUNT 366 a 66 c 104 g 351 t 214 others  
ORIGIN  
Query Match 1.7%; Score 55.6; DB 219; Length 1101;  
Best Local Similarity 48.1%; Pred. No. 0.0044;  
Matches 76; Conservative 24; Mismatches 58; Indels 0; Gaps 0;  
QY 3042 AATATTTGCTATAGGAGTGCCTTCCTGCAACGTGTAATAAATACATTTGTCCTCT 3101  
Db 971 WMAATTTTAAATACATTAATTTTATATACATTTTAAATTAACWMAAT 912  
QY 3102 TCAAAAAGATTTCTTTTAAATAATTTCTAGTATTAATTTTAAACAGTAAATA 3161  
Db 911 TTTAAACATTTTATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAAT 852  
QY 3162 ATTAAATAGCAATATCTATCTTATGACTCTTTT 3199  
Db 851 AAAMAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 814  
RESULT 3  
LOCUS CNS0167M 1201 bp DNA GSS 26-JUL-1999  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL106396.1 GI:5621701  
VERSION  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1201)  
AUTHORS Genoscope.  
TITLE Direct Submission

JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
COMMENT - Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila genome project (BDGP) - <http://www.edgp.ed.ac.uk>. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.  
FEATURES  
Source  
1..1201  
/organism="Drosophila melanogaster"  
/plasmid="pBeloBAC11"  
/db\_xref="taxon:7227"  
/clone\_lib="DrosBAC"  
/clone="BACN15M24"  
/note="end : T7"  
BASE COUNT 323 a 87 c 79 g 551 t 161 others  
ORIGIN  
Query Match 1.6%; Score 52.4; DB 219; Length 1201;  
Best Local Similarity 45.8%; Pred. No. 0.029;  
Matches 99; Conservative 19; Mismatches 98; Indels 0; Gaps 0;  
QY 2985 ATTCTGATTTGCTAAACGCCAGTCTCTAGGAGTTTCTCATCTTTAGGANA 3044  
Db 866 ATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 925  
QY 3045 TATTTGCTATAGGAGTGCCTTCCTGCAACGTGTAATAAATACATTTGTCCTCTCA 3104  
Db 926 AAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 985  
QY 3105 AAAAGATTTCTTTTAAATTTTCTAGTATTAATTTTAAACAGTAAATAAT 3164  
Db 986 TAATTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1045  
QY 3165 AATGACAAATATCTATCTTATGACTCTTTT 3200  
Db 1046 TATTAATAAATAATTTATTAATAAATTTTATTTT 1081  
RESULT 4  
LOCUS CNS020K7/c 1092 bp DNA GSS 12-MAY-2000  
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 222L11 of library G from Tetraodon nigroviridis, genomic survey sequence.  
ACCESSION AL175696.1 GI:7813753  
VERSION  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.  
REFERENCE 1 (bases 1 to 1092)  
AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissbach,J.  
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1092)  
AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissbach,J.  
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence  
JOURNAL Unpublished

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	55.8	1.7	1101	219	CNS00E01	AL069797 Drosophila
2	55.6	1.7	1101	219	CNS00E07	AL069797 Drosophila
3	52.4	1.6	1201	219	CNS0167M	AL106396 Drosophila
4	51.8	1.6	1092	220	CNS020K7	AL175696 Tetradon
5	51.4	1.6	1017	219	CNS0152J	AL106105 Drosophila
6	51.4	1.6	1101	219	CNS0039G	AL106330 Drosophila
7	50.8	1.6	723	220	CNS021AR	AL106652 Tetradon
8	50.6	1.6	1201	219	CNS016E0	AL106652 Drosophila
9	49.2	1.5	928	219	CNS00DKY	AL071665 Drosophila
10	48.6	1.5	637	220	CNS036CC	AL1029845 Tetradon
11	48.6	1.5	999	219	CNS016YI	AL107364 Drosophila
12	48.2	1.5	759	156	C23700	C23700 C23700 Dict
13	48.2	1.5	893	245	AZ53591	AZ53591 ENTRO88TR
14	48.2	1.5	964	219	CNS0173X	AL108567 Drosophila
15	48	1.5	1101	219	CNS0021J	AL061936 Drosophila
16	48	1.5	1101	219	CNS00DLI	AL068469 Drosophila
17	48	1.5	1200	219	CNS016CI	AL106572 Drosophila
18	47.8	1.5	828	219	CNS00A8E	AL054853 Drosophila
19	47.6	1.5	1101	219	CNS002HV	AL062726 Drosophila
20	47.4	1.5	516	156	C92840	C92840 C92840 Dict
21	47.4	1.5	580	156	C92114	C92114 C92114 Dict
22	47.4	1.5	836	219	CNS00K62	AL077933 Drosophila
23	47.2	1.5	1091	247	AZ669326	AZ669326 ENTL137TF
24	47.2	1.5	1101	219	CNS0177R	AL076697 Drosophila
25	47	1.5	1101	219	CNS0039G	AL063921 Drosophila
26	46.8	1.5	887	245	AZ547672	AZ547672 ENTEC23TF
27	46.8	1.5	923	248	AZ691829	AZ691829 ENTL168TR
28	46.8	1.5	1101	219	CNS00BEU	AL056687 Drosophila
29	46.8	1.5	1101	219	CNS017ZQ	AL108704 Drosophila
30	46.6	1.5	855	221	CNS04P5D	AL100874 Tetradon
31	46.2	1.4	756	239	AZ192623	AZ192623 SP_1021.B
32	46.2	1.4	822	219	CNS009CW	AL053618 Drosophila
33	46.2	1.4	1092	220	CNS020K7	AL175696 Tetradon
34	46.2	1.4	1101	219	CNS001FB	AL060732 Drosophila
35	46.2	1.4	1101	219	CNS00YZS	AL090742 Drosophila
36	46.2	1.4	1201	219	CNS016E0	AL106652 Drosophila
37	46	1.4	514	235	AO880124	AO880124 HS_4869.B
38	46	1.4	977	219	CNS017PG	AL108334 Drosophila
39	46	1.4	1001	219	CNS0155H	AL105023 Drosophila
40	45.8	1.4	533	32	AV680925	AV680925 Drosophila
41	45.8	1.4	1101	219	CNS008WC	AL052719 Drosophila
42	45.4	1.4	969	219	CNS006LP	AL065725 Drosophila
43	45.4	1.4	1101	219	CNS0039V	AL063936 Drosophila
44	45.4	1.4	1101	219	CNS0042W	AL055440 Drosophila
45	45.4	1.4	1101	219	CNS0100X	AL098379 Drosophila

## ALIGNMENTS

RESULT 1  
CNS00E01 1101 bp DNA GSS 04-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC:  
BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL069797  
VERSION AL069797.1  
KEYWORDS GI.4949738  
SOURCE GSS.

ORGANISM fruit fly.  
Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)

REFERENCE Genoscope.  
AUTHORS Direct Submission  
TITLE

## JOURNAL

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org/TheBDGP/Drosophila>  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mammosser in Pletier de Jong's Laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp. the same strain used for the BDGP's  
PI and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source  
1. 1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_id="RPCI-98"  
/clone="BACR29P15"  
/note="end : TET3"  
Location/Qualifiers

BASE COUNT 174 a 179 c 72 g 207 t 469 others  
ORIGIN

Query Match 1.7%; Score 55.8; DB 219; Length 1101;  
Best Local Similarity 25.7%; Pred. No. 0.0039;  
Matches 116; Conservative 127; Mismatches 209; Indels 0; Gaps 0;

QY 1 CAATGTGCAACAGACAGTACAGAAATGGATTCATAAACCACTTATATTA 60  
DB 561 CAAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 620  
QY 61 AATCTTACTTGGCTCATATTAATAGAAAACAGAGAGTCAAGATTAATCTTGAC 120  
DB 621 CACCAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 680  
QY 121 AGCTGTTTGCATCTTAACTTGAATTTTATTTTATTTTATTTTATTTTATTTTATTT 180  
DB 681 AATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 740  
QY 181 TCTCTAAACAAACAGATTCATGAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
DB 741 CATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 800  
QY 241 GTGTAGCTTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300  
DB 801 TTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 860  
QY 301 GATAGCTTTATTTGAAATATGATTCAGAACTTATTCCTCAAAACCTTACAGCACAACA 360  
DB 861 AATCAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 920  
QY 361 TATTTCTTACAGAGATGCTTCTTTACAGAGCTGGAAGAGCAGCTCCCTTATCTGAC 420  
DB 921 CACTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 980  
QY 421 AGTGTTTTAAAGCAACAGGAGCAATCTTAC 452  
DB 981 CMTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1012

## RESULT 2

CNS00E07/c 1101 bp DNA GSS 04-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC:  
BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL069440

117: gb\_est148:\*  
118: gb\_est149:\*  
119: gb\_est150:\*  
120: gb\_est151:\*  
121: gb\_est152:\*  
122: gb\_est153:\*  
123: gb\_est154:\*  
124: gb\_est155:\*  
125: gb\_est156:\*  
126: gb\_est157:\*  
127: gb\_est158:\*  
128: gb\_est159:\*  
129: gb\_est160:\*  
130: gb\_est161:\*  
131: gb\_est162:\*  
132: gb\_est163:\*  
133: gb\_est164:\*  
134: gb\_est165:\*  
135: gb\_est166:\*  
136: gb\_est167:\*  
137: gb\_est168:\*  
138: gb\_est169:\*  
139: gb\_est170:\*  
140: gb\_est171:\*  
141: gb\_est172:\*  
142: gb\_est173:\*  
143: gb\_est174:\*  
144: gb\_est175:\*  
145: gb\_est176:\*  
146: gb\_est177:\*  
147: gb\_est178:\*  
148: gb\_est179:\*  
149: gb\_est180:\*  
150: gb\_est181:\*  
151: gb\_est182:\*  
152: gb\_est183:\*  
153: gb\_est184:\*  
154: gb\_est185:\*  
155: gb\_est186:\*  
156: gb\_est187:\*  
157: gb\_est188:\*  
158: gb\_est189:\*  
159: gb\_est190:\*  
160: gb\_est191:\*  
161: gb\_est192:\*  
162: gb\_est193:\*  
163: gb\_est194:\*  
164: gb\_est195:\*  
165: gb\_est196:\*  
166: gb\_est197:\*  
167: gb\_est198:\*  
168: gb\_est199:\*  
169: gb\_est200:\*  
170: gb\_est201:\*  
171: gb\_est202:\*  
172: gb\_est203:\*  
173: gb\_est204:\*  
174: gb\_est205:\*  
175: gb\_est206:\*  
176: gb\_est207:\*  
177: gb\_est208:\*  
178: gb\_est209:\*  
179: gb\_est210:\*  
180: gb\_est211:\*  
181: gb\_est212:\*  
182: gb\_est213:\*  
183: gb\_est214:\*  
184: gb\_est215:\*  
185: gb\_est216:\*  
186: gb\_est217:\*  
187: gb\_est218:\*  
188: gb\_est219:\*  
189: gb\_est220:\*

190: gb\_est110:\*  
191: gb\_est111:\*  
192: gb\_hic:\*  
193: em\_gss\_fun:\*  
194: em\_gss\_hum1:\*  
195: em\_gss\_hum2:\*  
196: em\_gss\_hum3:\*  
197: em\_gss\_hum4:\*  
198: em\_gss\_hum5:\*  
199: em\_gss\_hum6:\*  
200: em\_gss\_hum7:\*  
201: em\_gss\_hum8:\*  
202: em\_gss\_hum9:\*  
203: em\_gss\_inv1:\*  
204: em\_gss\_inv2:\*  
205: em\_gss\_inv3:\*  
206: em\_gss\_other:\*  
207: em\_gss\_pln1:\*  
208: em\_gss\_pln2:\*  
209: em\_gss\_pro:\*  
210: em\_gss\_rod1:\*  
211: em\_gss\_rod2:\*  
212: em\_gss\_rod3:\*  
213: em\_gss\_rod4:\*  
214: em\_gss\_rod5:\*  
215: em\_gss\_vrt1:\*  
216: em\_gss\_vrt2:\*  
217: em\_gss\_vrt3:\*  
218: gb\_gss1:\*  
219: gb\_gss2:\*  
220: gb\_gss3:\*  
221: gb\_gss4:\*  
222: gb\_gss5:\*  
223: gb\_gss6:\*  
224: gb\_gss7:\*  
225: gb\_gss8:\*  
226: gb\_gss9:\*  
227: gb\_gss10:\*  
228: gb\_gss11:\*  
229: gb\_gss12:\*  
230: gb\_gss13:\*  
231: gb\_gss14:\*  
232: gb\_gss15:\*  
233: gb\_gss16:\*  
234: gb\_gss17:\*  
235: gb\_gss18:\*  
236: gb\_gss19:\*  
237: gb\_gss20:\*  
238: gb\_gss21:\*  
239: gb\_gss22:\*  
240: gb\_gss23:\*  
241: gb\_gss24:\*  
242: gb\_gss25:\*  
243: gb\_gss26:\*  
244: gb\_gss27:\*  
245: gb\_gss28:\*  
246: gb\_gss29:\*  
247: gb\_gss30:\*  
248: gb\_gss31:\*  
249: gb\_gss32:\*  
250: gb\_gss33:\*  
251: gb\_gss34:\*  
252: em\_gss\_inv4:\*  
253: em\_gss\_rod6:\*  
254: em\_gss\_rod7:\*  
255: em\_gss\_rod8:\*  
256: gb\_gss35:\*  
257: gb\_gss36:\*  
258: gb\_gss37:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2001, 12:45:42 ; Search time 10300.6 Seconds  
(without alignments)  
2936.646 Million cell updates/sec

Title: US-09-446-677B-1  
Perfect score: 3200  
Sequence: 1 CAATGTCGAGAGAGACTA.....TTCTATTGACTTCTTTT 3200

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estda:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estp11:*
73: em_estp12:*
74: em_estp13:*
75: em_estp14:*
76: em_estp15:*
77: em_estp16:*
78: em_estp17:*
79: em_estp18:*
80: em_estp19:*
81: em_estp110:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*
```



```

Db 478 TLSAEANISLNGLANVSLDGTNKAALTEADKNISLSTIALIDTEGSFYENHNK 537
QY 444 EDHSYDGIETL-DAGKDIYISADSRISINAVOSP---YGYGKWTINW--STDDKATVSM 497
Db 538 SASIYR-LIELTTAGANGTITIGALSTLTLOEPETHYGYGMOLSMANATSSKIGSIW 596
QY 498 AKOSENPTEAOEAPLPVNLWGSEFIDVPRFQNFTELGTGABYEKRFVYAGISNVLHRS 557
Db 597 TRGYIPSPERKSNLPLNLMGNFIDIRINSINOLIEKSSGEPFERELWLSGIANFEYRDS 656
QY 558 RENOKRFRHVSAGAVGASTRMPGGDTLSLGFQALFARKDYFMNTNFAKTYAGSLRLOH 617
Db 657 MTRHGFRIHSGYALGITATTPAEDQLTFAPCOLFARDRNHTGKNHDTGASLYFHH 716
QY 618 DASLVSVALLEGGLRELLPYVSKTLPSCSEFYGOLSYGHTDHRKTESLPPPTLTST 677
Db 717 TEGCLPDIANFLMGKATRAWVLSLSEISQIILPSFDAKFSYLHTDNHMK-----TYT 767
QY 678 DHT----SMGQYVWAGELGTTRAVENTSGRGEFFRETPPVKQAVYSRODSFVELGALS 733
Db 768 DMSITKGRNRNDAFCADLGASLPFV-ISVPYLKEVEPVPKQYIYAHQODFYERRAEGR 826
QY 734 DSDSHLYNLALPLGIKLEK--RFAEQYHYVAVMSPDYCRSNPKCTTLLSNQGSWKTK 791
Db 827 AENKSELINWEIPIGYTFERDSKSEKGYDILMYILDAVRRNPKQTSILIASDANMAY 886
QY 792 GSNLARQAGIYOASGFRSLGAAAEELGNGFEMRGSSRSRYNDAGSKIFE 841
Db 887 GTNLARQGSVRAANHFQVNPHEIFGQFAFEVRSSRMYNTNLGSKFCF 936

RESULT 15
C86546
polymorphic outer membrane protein G family [imported] - Chlamydothia pneumoniae (strain
C:Species: Chlamydothia pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: C86546
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: AB6491; MUID:20330349
A:Accession: C86546
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-936 <STO>
A:Cross-references: GB:BA000008; NID:98978817; PIDN:BA98653.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: pmp_7

```

Query Match 30.9%; Score 1328.5; DB 2; Length 936;  
 Best Local Similarity 32.5%; Pred. No. 6.8e-69;  
 Matches 309; Conservative 158; Mismatches 360; Indels 123; Gaps 14;

```

QY 1 MKIPIRFLILSLVPTLSNMLGAATTEELASNSPDTTSTTSFSSKTSATDGTNYVF 60
Db 1 MMSVSWLFFSSIPFSSLSIYAAEVLTD-SSNNSYDGSNGTTFYFTTDAAGTYSL 59
QY 61 KRSYVLENPKTGEQOSCFKNDAAAGDLNLFGGFSTFESNIDATTAAGAAISSEAN 120
Db 60 LADVSFQNGALGIPLAGSCFLE--AGGDLTFQGNQHALKFAFINAGSAGTVASTSAAD 117
QY 121 KTVTLSGFALSFLKPASTVF-NGLGAINVKNLSLLDNKVLIODNFSTGDGAIN-- 177
Db 118 KMLLNDPFRSLIISCPLLSLPTGOCALKSVGNLSLGNISQIIFTQNFSDNGYINTK 177
QY 178 -----CAGSLKIAN----- 186
Db 178 NFLLSTQFASFSRQAFATGKGQGVVYATGTTIENSFGIVSFQNLAKSGGALYSTD 237
QY 187 -----NKSLSFGNSSSTRGAIH 205

```

Search completed: October 2, 2001, 03:29:09  
 Job time: 1326 sec

```

Db 238 NCSITDNFOVIFDGSNWEAAOAGALICCTTDTKVTTLTGKNLSFTNNALTYGALIS 297
QY 206 TKNLTLSSGGETLFGONTPTAAGK--GGAILADSGTSLISGDSGDIIEGNTIGATGT 263
Db 298 GLKVTISAGGPTLFQSNISGSSAGGGAATINASGELALNTSDDITFFNNQYTNGST 357
QY 264 VSHAIDLQTSKAITALRAAQHTTYFDPTVTYSTVADALNINSPDTGNREYTGTL 323
Db 358 STRNMNITDFAKYSIRATQOSIYFYDPTINPGTAATDILNLNLADANSEIEYGAI 417
QY 324 VSGEKLTEAEKDEKRNKSKLLQNVAFKNGTVVLKGDVYLSANGFSQANSKLINDLT 383
Db 418 VESGELSTPEKAIAANYVTIROPVILARGDLVLDGYTVFEKDLTQSPGRNLIIMDGT 477
QY 384 SILVANTESIELNLEINIDSLNNGKRIKLSAATPAOKDIRIDPRVYLATIDESFYONGFIN 443
Db 478 TLSAEANISLNGLANVSLDGTNKAALTEADKNISLSTIALIDTEGSFYENHNK 537
QY 444 EDHSYDGIETL-DAGKDIYISADSRISINAVOSP---YGYGKWTINW--STDDKATVSM 497
Db 538 SASIYR-LIELTTAGANGTITIGALSTLTLOEPETHYGYGMOLSMANATSSKIGSIW 596
QY 498 AKOSENPTEAOEAPLPVNLWGSEFIDVPRFQNFTELGTGABYEKRFVYAGISNVLHRS 557
Db 597 TRGYIPSPERKSNLPLNLMGNFIDIRINSINOLIEKSSGEPFERELWLSGIANFEYRDS 656
QY 558 RENOKRFRHVSAGAVGASTRMPGGDTLSLGFQALFARKDYFMNTNFAKTYAGSLRLOH 617
Db 657 MTRHGFRIHSGYALGITATTPAEDQLTFAPCOLFARDRNHTGKNHDTGASLYFHH 716
QY 618 DASLVSVALLEGGLRELLPYVSKTLPSCSEFYGOLSYGHTDHRKTESLPPPTLTST 677
Db 717 TEGCLPDIANFLMGKATRAWVLSLSEISQIILPSFDAKFSYLHTDNHMK-----TYT 767
QY 678 DHT----SMGQYVWAGELGTTRAVENTSGRGEFFRETPPVKQAVYSRODSFVELGALS 733
Db 768 DMSITKGRNRNDAFCADLGASLPFV-ISVPYLKEVEPVPKQYIYAHQODFYERRAEGR 826
QY 734 DSDSHLYNLALPLGIKLEK--RFAEQYHYVAVMSPDYCRSNPKCTTLLSNQGSWKTK 791
Db 827 AENKSELINWEIPIGYTFERDSKSEKGYDILMYILDAVRRNPKQTSILIASDANMAY 886
QY 792 GSNLARQAGIYOASGFRSLGAAAEELGNGFEMRGSSRSRYNDAGSKIFE 841
Db 887 GTNLARQGSVRAANHFQVNPHEIFGQFAFEVRSSRMYNTNLGSKFCF 936

```



```

Db 172 GSSISLSNPLUFAKKNKATQKGALYSTGCTITNNTLNSASFSENFANNGALYTEAS 231
Qy 170 -----TGDGAINCAGS-----LKIANNSLSFTIGSSSTRGCAI 204
Db 232 SFISSNKAISFNNINNSVATPSATGCAIYCSTSPAKPVLTLLSDNGELNFTIGNTALISGCAI 291
Qy 205 HTKNLTSSGGETTFQGTAL-PTAAGKGAIALDSCGLTISGDSGDIIIFGENT--GAT 261
Db 292 YTDNLVSSGGPTLFKNNSAIDTAAPLGGAIALDSCGLSLSALGDDITFEONTVYKAS 351
Qy 262 G--TVSHSIDLG--TSAKITAPLAAOGHTIYFDPITYGTSVADALINSPDGDKE 318
Db 352 SSQOTTRNSINIGTNAKIYQVLASQGTITFYDPTITSTIALSDALNGLPDLAGPA 411
Qy 319 YTGTFVSGEKLTEAEAKDEKNRTSKLLQNYAFKNGTVLKGDDVYLSANGFESQDANSKLI 378
Db 412 YOGTFVSGEKLTEAEAEADNLKSTIQOPLTLAGGQSLKSGVTLVAKSFSQSPGFTLL 471
Qy 379 MDLGTSLVANTESIELTULEINIDSLRNGKIKLSAATAQDIRIDRVYVLAISDESYQ 438
Db 472 MDAQTTL-ETADGITINNVLNVSDLSKETKATLKAQASQVTLSSGLSLVDPSCANYE 530
Qy 439 NGFLNEDHSYGI-LELDAGKDIY---SADSRSINAVOSPVGQKWTIYW---STDD 490
Db 531 DVSNNRPVEFSCUHLITDADPANIHITDLAADPLENPH--WGIOGNALMSHOETAKRS 588
Qy 491 KKATVSMAKGSFNPTAEQADLPVNLWGSFIDVAPFQNFTELGETGAPERYKRFVWAGIS 550
Db 589 KAATLTWTKTYNPNPBERGTLVANTLWGSFVDVASIQVLATKVRQSETRGIMCEGIS 648
Qy 551 NVLHRSGENORRRHVSAGAVVAGASTPMPCGDILSLCFADLPARDKQYFMNTNFAKTYA 610
Db 649 NFFHKDSTIKKRGRIHSIAGVYVATYTTLASDNLITTAFCOLFGRDHPFINKNASAYA 708
Qy 611 GSLRLQHDASISVYVSLILGEGRLREILPYVSKTLPISFGOLSYGTHDRMKTRESLP 670
Db 709 ASLHLQHLATL-SSPSLIR-----YLPSESSEQVLPFLNAQISITYISKNTMKTYTQA 759
Qy 671 PPPTLSDTSDTSMGWYVWAGELGTIRAVAVENTSGRGFFRYTFPVKQAVYSRODSFEVLG- 729
Db 760 P-----KGESSWYNDGCLATELASSLPHTLALHHEGLFHAHPPIKVEASYIHQDSKENT 814
Qy 730 AISRDESDSHLYNLAIPLGIKLEK--RPAEOYHHVAMYSPOVCRSNPKCTTTLSNCGS 787
Db 815 TLVFSFGDGLINVSVPIGITFEERFSRNERASYEATVIYVADVYKKNPDCYTTALLINNTS 874
Qy 788 WKTGGSNARAGIYVQASGFRSLGAAALFPNGFGEWGSRSRYVNDGSKITF 841
Db 875 WKTGTNLISRQAGIRAGITFAFSPNLEVTNLSMEIKGSSRSRYVADAGKRFQ 928

```

RESULT 12  
B72077  
polymorphic membrane protein G family CP0306 [Imported] - Chlamydophilia pneumoniae (Streptococcus pneumoniae)  
C:Species: Chlamydophilia pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: B72077; B81592  
R:Kalmun, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Ollinger, L.; Grimwood, J.; et al.  
Nature Genet. 21: 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: B72077  
A:Molecule type: DNA  
A:Residues: 1-928 <ARN>  
A:Cross-references: GB:AE001628; GB:AE001363; NID:g4376730; PIDN:AADI8591.1; PID:g4376730  
A:Experimental source: Strain CWL029  
R:Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.L.  
Nucleic Acids Res. 28: 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis Moyn and Chlamydia pneumoniae AR39.  
A:Reference number: AB1500; MUID:20150255  
A:Accession: B81592  
A:Molecule type: DNA

A;Residues: 1-928 <RFA>  
A;Cross-references: GB:AE002199; GB:AE002161; NID:g7189226; PIDN:AAF38163.1; PID:g718  
A;Experimental source: strain AR39, HL cells  
C;Genetics:  
A;Gene: pmp\_9; CFP0306

Query Match	32.0%;	Score 1376.5;	DB 2;	Length 928;
Best Local Similarity	-36.6%;	Pred. NO. 1.1e-71;		
Matches 349;	Conservative 125;	Mismatches 341;	Indels 139;	Gaps 23;

```

OY      1 MKRILARELLI--SLVPTLMSMLNGLAATTE--ELISANSFDOTTTSTSPSSKTSATGTN 57
Db      1 MKSSLHMFLLISSLALPLSLNFSAFAAFAVEINLCPTNFSFGPGTYTTPPAQTNA--DGTI 58
OY      58 YVFDSVVIENVPKTGETOSTCEPKNDAAAGDLNLFLOGEFSFTESNIDATTAASCAIGSE 117
Db      59 YNLGADVSIINA--GSPALNATSCFKE--TGNLSFGFHGAVFLQINIDA--GANCFTTNT 113
OY      118 AANKTVTLGFSALSFLKSPASTVINGICALINVKNTSLDNDKVLITQDNFS----- 165
Db      114 AANKLSPSGFSLSLIQT--TNATYTGAIKISGACISIGSNYSCYGFQNFSDNDGALQ 177
OY      170 ----- 165

```

Db 172 GSSISLSLNPNTFAKNKATQKGALYSTGCITINNTLSASFSENTAANNNGAIYTEAS 231

QY 170 -----TG D G G A I N C A G S -----L K I A N K S L S F I G N S S T R G G A I 204

Db 232 SEISSNKRAISEINNSVATATSGATGYCSSTAPKPVITLSDNGEI NEIGNTATSGCAT 291

COE UNIVERSITY POLYMER DEPARTMENT OF CHEMISTRY

[illegible]

DD 434 11DNLE55GGF1LF KNNBSA1D1A4F LGGAI1A1AUB5G5L5L5AUGD11FEGN1VVA6AS 331

262 G--TVSHSAIDLG-TSAKITALRAAQHTIFYDDPITVTGSTSADALNINSPDTGDNKE 318

Db 352 SSQTTTNSINIGNTNAKIVQLRASQGNITYFYDPITTSITAAALSDALNLNGPDLAGNPA 411

QY 319 YTGTVFSGEKLTEAEAKDEKNRTSKLLQNVAFKNGTVVLKGDVLSANGFSQDANSKLI 378

Db 412 YOGTIVFSGEKLSEAAEAADNLKSTIOOPLTLAGGOLSKSGVTLVAKSFSOSPSTLL 471

379 MDIGSTVANTESTETNIETINDSLRNGKKTKI.SAATAOKDTRIDBPVYIATSDESEFYQ 438

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible]

439 NGF UNEDHSIDGI -LEDDAGNDVI -SADSKSINAVQSFIGIQKMI INW---SIDD 430  
:  
:  
:

Db 531 DVSNNPQVFSCLLTADDPANIHITDPLADPLEKNPIH--WG YQGNWALSMQEDTATKS 588

491 KKATVSWAKQSFNPTAEQEARPLVPNLLWGSFIDVRPFQNFIELGTGEGAPYEKRFWAGIS 550

Db 589 KATLTWTKTGYNPERRGTLVANTLWGSFVDVRSIQQLVATKVRQSQETRGICGIS 648

QY 551 NVLHRSGRNQKFRHVSGAVVGASTRMPGGDTLSLGEAQLFARDKDYFMNTNFAKTYA 6100

```
db      649 NEHKDSTKINKEBHISAGYVGATTTI.ASDNL.TTAAECOL.EGKDBDHETINKBRASAYA 708
```

[illegible]

10  
9  
8  
7  
6  
5  
4  
3  
2  
1

00 /03 402H44H4444 23F3244 ----- 14F02E2E4F4444444444444444 /03

QY 6/1 PPPTLSTIDHTSMGGYVWAGELSTRVAVENTSGRGFEFREYTPFKVQAVSKQDSFVELG- /22

Db 760 P-----KGESSWYNDGCALELASSLPHTALSHEGLFHAFFPIKVEASYIHQDSFERNT 814

QY 730 AISRDESDSHLYNLAIPLGKLEK--RFAEQYYHVAMYS PDVCRSNPKCTTLLSNQGS 787

Db 815 TLVRSFDGDLINVSPIGITFERFSRNERASYEATVIYVADYRKPNPDCITALLINNTS 874

0V 788 WKTGCSNI.ABOAGTVOASGFRSIGAAAFI.EGNEGEFEWBGSSRPSYNDAGSKTKE 841

[illegible]





```
QY 379 MDLGTSLVANTSEITELNLEINIDSLRNGKKIKLSAATAOKDIRIDRPVLAIDSESYQ 438
D 473 MOPGTKLKADTEALSLTKLVYDLSALEGNKSVSIETAGANKTITLSPVLYRQDSSGNRYE 532
QY 439 NGFLNEDHSYDGL--ELDAGKDIVISADSRISNAVOSP--YGYQKWTINW--STDDK 491
D 533 SHITNOAFTOPLVFTATASDIYI--DALITSPVOTPEPHYGYQGHMEATWADTSTAK 590
QY 492 KATVSMKOSFNPTAEQAPLVPLNLGSEFIDVRPFQNFIELEGABYERKRVAGISN 551
D 591 SGTMTWTTGTGNPNERASVYVPSLWASFIDITLQOIMTSQANSIYQOGLWASGTAN 650
QY 552 VLHSGRENOGRFPHVSGAVGASTRPGDITSLGPAOLFARDKOVEMTNAKTATAG 611
D 651 FPHDKSGTNOAFPHKSGYIVGSGADESENFISVAFQGLFGKDKDLFIENSHNTLA 710
QY 612 SLRQHDASLVSYSILLGEGREILLPYVS-----KTLPCSFYQGLSYGHTDHRM 663
D 711 SLVYQHRAFL-----GG-----LPMSFGSITDMLKDIPILINAOLSYSTYTKNDM 755
QY 664 KTESLPPPPPLSTDHNSMGYVWAGELGTRVAVENTSGRGEFREYTPFVQVAVSRQD 723
D 756 DTRTYSTPEA-----QGSWTNNSGALLEGSLALYLPEKAEFFQGYPEPLTKFOAVSRQ 810
QY 724 SFVELGAISRDFSDSHLYNLALPLGIRLEKRFARQ--YHYVNAVYSPDYCRSNPKCTTTL 781
D 811 NFKSGAEARAFDDGDLVNCISIPVIGIRLEKISEDEKNFELISLAYIGDYRYKRNRSRTSL 870
QY 782 LSNQSGSKTKGSNLARQAGIYQVQASGFRSLGAAAEFLGNGFEGEMGSSRSYVNDAGSKIRF 841
D 871 MVSGASWTSILCKNLARQAFILASAGSHLTLSPHVELSGEAAVELRGSNHIYVNDGGLRYSF 930

RESULT 7
D72078
polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: D72078
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: D72078
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-930 <ARN>
A:Cross-references: GB:AE001627; GB:AE001363; NID:g4376721; PIDN:AAD18590.1; PID:g437672
C:Experimental source: strain CWL029
C:Genetics:
A:Gene: pmp_8

Query Match 32.9%; Score 1413.5; DB 2; Length 930;
Best Local Similarity 37.0%; Pred. No. 8.4e-74;
Matches 355; Conservative 124; Mismatches 332; Indels 149; Gaps 20;
```

```
QY 170 -----TGDDGAINC-----AGSLKIANKSLSPFGNSSRPG 201
D 234 FEASSSTONSLFPGTATDAGKGAITCETGTEPILITIGNSLTPAEMSSVTQ 293
QY 202 GAITKNTLSSGGETLFOGN-TAPTAAGKGAITADSGTSLSCSDGDIIFEGNTIGA 260
D 294 GAIAGHIDLSAAPTLEFSNNRCNTAAGKGAITADSGTSLSANOGDITFGNTLS 353
QY 261 TG--TVSHSIDCTSAKTALRAAQCHTTFYDPIYVYGSTVADALNINSPDTGDK 318
D 354 TSAPTSTNAIYILSSAKITNLRAAQOSIIFYDPI-ASNTGASDVLTINOPDPSNPLD 412
QY 319 YTGTVSGEGLTEAEKDEKNRFTSKLONAFNGVYLVKGDVVLANSNGSODANSKLI 378
D 413 YSGTVISGEKLSADENKADNFTSILKOPALASGTLALGNVELDNGFTQEGSTLL 472
QY 379 MDLGTSLVANTSEITELNLEINIDSLRNGKKIKLSAATAOKDIRIDRPVLAIDSESYQ 438
D 473 MOPGTKLKADTEALSLTKLVYDLSALEGNKSVSIETAGANKTITLSPVLYRQDSSGNRYE 532
QY 439 NGFLNEDHSYDGL--ELDAGKDIVISADSRISNAVOSP--YGYQKWTINW--STDDK 491
D 533 SHITNOAFTOPLVFTATASDIYI--DALITSPVOTPEPHYGYQGHMEATWADTSTAK 590
QY 492 KATVSMKOSFNPTAEQAPLVPLNLGSEFIDVRPFQNFIELEGABYERKRVAGISN 551
D 591 SGTMTWTTGTGNPNERASVYVPSLWASFIDITLQOIMTSQANSIYQOGLWASGTAN 650
QY 552 VLHSGRENOGRFPHVSGAVGASTRPGDITSLGPAOLFARDKOVEMTNAKTATAG 611
D 651 FPHDKSGTNOAFPHKSGYIVGSGADESENFISVAFQGLFGKDKDLFIENSHNTLA 710
QY 612 SLRQHDASLVSYSILLGEGREILLPYVS-----KTLPCSFYQGLSYGHTDHRM 663
D 711 SLVYQHRAFL-----GG-----LPMSFGSITDMLKDIPILINAOLSYSTYTKNDM 755
QY 664 KTESLPPPPPLSTDHNSMGYVWAGELGTRVAVENTSGRGEFREYTPFVQVAVSRQD 723
D 756 DTRTYSTPEA-----QGSWTNNSGALLEGSLALYLPEKAEFFQGYPEPLTKFOAVSRQ 810
QY 724 SFVELGAISRDFSDSHLYNLALPLGIRLEKRFARQ--YHYVNAVYSPDYCRSNPKCTTTL 781
D 811 NFKSGAEARAFDDGDLVNCISIPVIGIRLEKISEDEKNFELISLAYIGDYRYKRNRSRTSL 870
QY 782 LSNQSGSKTKGSNLARQAGIYQVQASGFRSLGAAAEFLGNGFEGEMGSSRSYVNDAGSKIRF 841
D 871 MVSGASWTSILCKNLARQAFILASAGSHLTLSPHVELSGEAAVELRGSNHIYVNDGGLRYSF 930

RESULT 8
A81591
polymorphic membrane protein G family CP0307 [imported] - Chlamydia pneumoniae (s
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: A81591
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heldelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: A81591
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-930 <REA>
A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AAF38164.1; PID:g718
C:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0307

Query Match 32.9%; Score 1413.5; DB 2; Length 930;
Best Local Similarity 37.0%; Pred. No. 8.4e-74;
```





```
QY 121 KTVLTSFSALEFLKSPASTVNTGLGAINVKGNLSLDNDKVLIDQDNFSTGGGAINCAG 180
    ||||| ||| ||||| ||| : : : : : ||||| |||||
Db 112 KNLTFSGFSLSPDSSPTVTTGGTLLSAGVNLNIRKLVAAGNFSTADGAIKAGS 171
QY 181 ----- 180
Db 172 FLTGTSGLDALFSNNSSSTKGAIAATTAGARIANNNGYVRFSLNASTSGAIDDEGTSI 231
QY 181 ----- SLKTAANKSLSEFIGNSSSTGCAIHTKNTL 210
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 232 LSNKKFLYFEGNAKKTGCAICNTKASGSPELLISNNKTLIFASNVAETSGCAIHAHAKLA 291
QY 211 LSSGFTLFGONTAPTAAGKGAIAIADSGTLLSISGSDIIFEGNPTI---GATGYVSHS 267
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 292 LSSGFTFELRNNAVSAATPKGAISIDAGSELISAEFGNTTFVRNLTGTTGSDTPKRN 351
QY 268 AIDGTSAKITLALAAAGHTIYFYDPTTVGTSVADALNINSPDGDNKEEYGTIYFSG 327
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 352 AINNGSGKFTLBAKNNHTIIFYDPTTSEGTSS--DVLKINNSAGALNPYOGTILFSG 409
QY 328 EKLTEAKDEKNTSKLQNVAFKNGTVVILKGDVLSANGFSDANSKLIIMDGTSLVA 387
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 410 EYTLADELKAADNLSKSFQPVSLSGCKLLQKGVTLSESTFSQEAAGSLGMDGTLLST 469
QY 388 NTESELTNLEINIDSLRNKIKLSAATAOKDIRIDRPVYLAISDSFYONGFLNEDHS 447
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 470 TAGSITITNLGINVDSGLKQPVSLTAKGASNKYVSGKLLIDIEGNIYESHMFSDOL 529
QY 448 YDGLIEDACKDVIYASDSRSINAV-----QSPYGOGKWTIMNSTD---DKKATVSWA 498
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 530 F-SLTKITVDADVDNTNDISLIPVPAEDPNSEYFGOGNNVMTTDTATNTKEATATWT 588
QY 499 KOSNPTAEDEAPLYPMLNGSFIDVPPQNFIELGTGEGAYEKRFVAGISNVLHSGR 558
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 589 KTGVPSPERKSALVCNTLNGVFTDILSDQVLGATGHEHKGQFVWSMTNPLHKTGD 648
QY 559 ENOKRFHVSGAAGVASTRMPGDDTLISGFAOLFARDKDFMNTNPAKTYAGSLRLQHD 618
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 649 ENKRGFHTSGGYIYGSAHTPKDLETFPAFCHLFARDKCFIAHNNSRITGGTLFPRKS 708
QY 619 ASLYSVSILLBEGGLREILLPYVSKTLPCSEFYGQLSYGHTDHRMKE--SLPPPTPLS 676
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 709 HTLQPOMYLRLGRKFSESAIEKFPREIPLADVOVSFSHSDNMEHTYSLP----- 761
QY 677 TDHRTSMGYYWAGELGRVAVENTSGRFFREYTPPVKQAVYISRODSFVELGAIISDF 736
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 762 ESEGSWNECIAGIGIDLFPVLSNPHPLKTFIPQKVEVYVYSONSFESSSDGGRGFS 821
QY 737 DSHLYNLAIPLGIR-LKKRPAEOY-YHVVAAMYSPDVCRSNPKCTTLLSNGSKRTGSGN 794
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 822 IGRLLNLSIPGAKFYVGDIGDSTYTLDSGFVSDVYRNPNQSTATLVMSPDSKIRIGGN 881
QY 795 LARQAGIVQASGFRSLGAALFEGNFGFEMWGRSSRYNVDAGSKIRF 841
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 882 LSRQAFLLRGSNNVYVNSNCELFGHYAMELRGSSRYNVDVGTKLRF 928

RESULT 4
D72077
polymorphic outer membrane protein g family - Chlamydomophila pneumoniae (strain CWL029)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: D72077
R:Rahman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; NCID:99206606
A:Accession: D72077
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-928 <ARN>
A:Cross-references: GB:AE001628; GB:AE001363; NID:g4376730; PID:AA018593.1; PID:g437673
```

```
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: pmp_11

Query Match 33.3%; Score 1429.5; DB 2; Length 928;
Best Local Similarity 35.8%; Pred. No. 1e-74;
Matches 339; Conservative 147; Mismatches 336; Indels 125; Gaps 15;

QY 1 MKTIPRLILSVPTLSNLSLGAATTELSASNSPDGTTSTTSSKTSATDCTNYV 60
    || : : : || : : : || : : : || : : : || : : : ||
Db 1 MKTIPVVLVSVLAFS-CHLOSLANELELSPDDSFNGNIDSGFTFTKTS----TTFYSL 55
QY 61 KDSVIENVPKGTGOSTSCPKNDAAGDLEFGGGSFSPFNSDATATAGAAIGSAAAN 120
    || : : : || : : : || : : : || : : : || : : : ||
Db 56 TGDVFEYE-PEGKTPPLSDSCPKQ--TTDNLTFELNGHSLTFEGFDAGTHGAA-ASTTAN 111
QY 121 KTVLTSFSALEFLKSPASTVNTGLGAINVKGNLSLDNDKVLIDQDNFSTGGGAINCAG 180
    ||||| ||| ||||| ||| : : : : : ||||| |||||
Db 112 KNLTFSGFSLSPDSSPTVTTGGTLLSAGVNLNIRKLVAAGNFSTADGAIKAGS 171
QY 181 ----- 180
Db 172 FLTGTSGLDALFSNNSSSTKGAIAATTAGARIANNNGYVRFSLNASTSGAIDDEGTSI 231
QY 181 ----- SLKTAANKSLSEFIGNSSSTGCAIHTKNTL 210
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 232 LSNKKFLYFEGNAKKTGCAICNTKASGSPELLISNNKTLIFASNVAETSGCAIHAHAKLA 291
QY 211 LSSGFTLFGONTAPTAAGKGAIAIADSGTLLSISGSDIIFEGNPTI---GATGYVSHS 267
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 292 LSSGFTFELRNNAVSAATPKGAISIDAGSELISAEFGNTTFVRNLTGTTGSDTPKRN 351
QY 268 AIDGTSAKITLALAAAGHTIYFYDPTTVGTSVADALNINSPDGDNKEEYGTIYFSG 327
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 352 AINNGSGKFTLBAKNNHTIIFYDPTTSEGTSS--DVLKINNSAGALNPYOGTILFSG 409
QY 328 EKLTEAKDEKNTSKLQNVAFKNGTVVILKGDVLSANGFSDANSKLIIMDGTSLVA 387
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 410 EYTLADELKAADNLSKSFQPVSLSGCKLLQKGVTLSESTFSQEAAGSLGMDGTLLST 469
QY 388 NTESELTNLEINIDSLRNKIKLSAATAOKDIRIDRPVYLAISDSFYONGFLNEDHS 447
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 470 TAGSITITNLGINVDSGLKQPVSLTAKGASNKYVSGKLLIDIEGNIYESHMFSDOL 529
QY 448 YDGLIEDACKDVIYASDSRSINAV-----QSPYGOGKWTIMNSTD---DKKATVSWA 498
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 530 F-SLTKITVDADVDNTNDISLIPVPAEDPNSEYFGOGNNVMTTDTATNTKEATATWT 588
QY 499 KOSNPTAEDEAPLYPMLNGSFIDVPPQNFIELGTGEGAYEKRFVAGISNVLHSGR 558
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 589 KTGVPSPERKSALVCNTLNGVFTDILSDQVLGATGHEHKGQFVWSMTNPLHKTGD 648
QY 559 ENOKRFHVSGAAGVASTRMPGDDTLISGFAOLFARDKDFMNTNPAKTYAGSLRLQHD 618
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 649 ENKRGFHTSGGYIYGSAHTPKDLETFPAFCHLFARDKCFIAHNNSRITGGTLFPRKS 708
QY 619 ASLYSVSILLBEGGLREILLPYVSKTLPCSEFYGQLSYGHTDHRMKE--SLPPPTPLS 676
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 709 HTLQPOMYLRLGRKFSESAIEKFPREIPLADVOVSFSHSDNMEHTYSLP----- 761
QY 677 TDHRTSMGYYWAGELGRVAVENTSGRFFREYTPPVKQAVYISRODSFVELGAIISDF 736
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 762 ESEGSWNECIAGIGIDLFPVLSNPHPLKTFIPQKVEVYVYSONSFESSSDGGRGFS 821
QY 737 DSHLYNLAIPLGIR-LKKRPAEOY-YHVVAAMYSPDVCRSNPKCTTLLSNGSKRTGSGN 794
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 822 IGRLLNLSIPGAKFYVGDIGDSTYTLDSGFVSDVYRNPNQSTATLVMSPDSKIRIGGN 881
QY 795 LARQAGIVQASGFRSLGAALFEGNFGFEMWGRSSRYNVDAGSKIRF 841
    ||||| ||| : : : : : ||||| ||| : : : : :
Db 882 LSRQAFLLRGSNNVYVNSNCELFGHYAMELRGSSRYNVDVGTKLRF 928
```

```

QY 301 SVADALININSPDGNKKEYTGTIVFSGEKLTEAEAKDEKRTSKLLQNYAFKNGTVLKG 360
DB 301 SVADALININSPDGNKKEYTGTIVFSGEKLTEAEAKDEKRTSKLLQNYAFKNGTVLKG 360
QY 361 DVLVLSANGFSQDANSKLLMDLCTSLVANTESIETLNLEINIDSLRNGKKIKLSAATAQKD 420
DB 361 DVLVLSANGFSQDANSKLLMDLCTSLVANTESIETLNLEINIDSLRNGKKIKLSAATAQKD 420
QY 421 IIRIDRPVVAISDESFYONGFLNEHSDYDGLLELDAGKDIVISAOSRSINAVOSPYGYG 480
DB 421 IIRIDRPVVAISDESFYONGFLNEHSDYDGLLELDAGKDIVISAOSRSIDAVOSPYGYG 480
QY 481 KWTIMSTDDKATYVMAKOSFNPTEAEAPLVNLMGSEFIDVRFQNFIELGTEGARY 540
DB 481 KWTIMSTDDKATYVMAKOSFNPTEAEAPLVNLMGSEFIDVRFQNFIELGTEGARY 540
QY 541 EKRFVWAGISNVLSHSGRENORRFRHVSAGAVGASTRMPGCDTSLGFAQLFARDKDYF 600
DB 541 EKRFVWAGISNVLSHSGRENORRFRHVSAGAVGASTRMPGCDTSLGFAQLFARDKDYF 600
QY 601 MNTNFAKTYAGSLRLOHDSILYSVSIILGEGLEIRELLPYVSKTLPCEFYGOLSTGHD 660
DB 601 MNTNFAKTYAGSLRLOHDSILYSVSIILGEGLEIRELLPYVSKTLPCEFYGOLSTGHD 660
QY 661 HRMKTESLPPPPPTLSTDHTSMGCVYVWAGELCTRYAVENTSGRGCFREXTPEVKYQAVYS 720
DB 661 HRMKTESLPPPPPTLSTDHTSMGCVYVWAGELCTRYAVENTSGRGCFREXTPEVKYQAVYS 720
QY 721 RODSVEVEIGAISRDSDSHLYLALPLGKLEKRFABOYHYVAMSPDVCNSPKCTTT 780
DB 721 RODSVEVEIGAISRDSDSHLYLALPLGKLEKRFABOYHYVAMSPDVCNSPKCTTT 780
QY 781 ILSNOSGSMKTKGSNLARAGIVQASGFSRSLGAAELFGNFGFEMHSGSSRYVWDAGSKIK 840
DB 781 ILSNOSGSMKTKGSNLARAGIVQASGFSRSLGAAELFGNFGFEMHSGSSRYVWDAGSKIK 840
QY 841 F 841
DB 841 F 841

RESULT 2
E86492
polymorphic outer membrane protein G family [imported] - Chlamydia pneumoniae (strat
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: E86492
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; I
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: E86492
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-712 <STO>
A:Cross-references: GB:BA000008; NID:98978386; PIDN:BA98223.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: pmp_2_1

```

```

Query Match 79.0%; Score 3394; DB 2; Length 712;
Best Local Similarity 99.7%; Pred. No. 2.9e-188;
Matches 670; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 MKIPRFLISLVPTLMSNLGAATTEELASNSFDGTTSTTSKTSATDGTNYF 60
DB 1 MKIPRFLISLVPTLMSNLGAATTEELASNSFDGTTSTTSKTSATDGTNYF 60
QY 61 KDSVVIENPKTGEGOSTSCFKNDAAGDLNLFGGGFSFTFENIDATTAAGAAIGSEAN 120
DB 61 KDSVVIENPKTGEGOSTSCFKNDAAGDLNLFGGGFSFTFENIDATTAAGAAIGSEAN 120

```

```

QY 121 KTYTSLGSFALSFLKSPASTVINGLGAIVNKGNSLLDNDKYLIDNFTSGDGAINCAG 180
DB 121 KTYTSLGSFALSFLKSPASTVINGLGAIVNKGNSLLDNDKYLIDNFTSGDGAINCAG 180
QY 181 SLKIANKSLSEIGNSSSTRGGAHTRKNTLSSGGETLFQGNAPTAAKGGAIAIADSG 240
DB 181 SLKIANKSLSEIGNSSSTRGGAHTRKNTLSSGGETLFQGNAPTAAKGGAIAIADSG 240
QY 241 TISISGSDGDIIFEGNTICATGTVSHSAIDLTGSAKITLRAAQGHITFYDPTTGST 300
DB 241 TISISGSDGDIIFEGNTICATGTVSHSAIDLTGSAKITLRAAQGHITFYDPTTGST 300
QY 301 SVADALININSPDGNKKEYTGTIVFSGEKLTEAEAKDEKRTSKLLQNYAFKNGTVLKG 360
DB 301 SVADALININSPDGNKKEYTGTIVFSGEKLTEAEAKDEKRTSKLLQNYAFKNGTVLKG 360
QY 361 DVLVLSANGFSQDANSKLLMDLCTSLVANTESIETLNLEINIDSLRNGKKIKLSAATAQKD 420
DB 361 DVLVLSANGFSQDANSKLLMDLCTSLVANTESIETLNLEINIDSLRNGKKIKLSAATAQKD 420
QY 421 IIRIDRPVVAISDESFYONGFLNEHSDYDGLLELDAGKDIVISAOSRSINAVOSPYGYG 480
DB 421 IIRIDRPVVAISDESFYONGFLNEHSDYDGLLELDAGKDIVISAOSRSIDAVOSPYGYG 480
QY 481 KWTIMSTDDKATYVMAKOSFNPTEAEAPLVNLMGSEFIDVRFQNFIELGTEGARY 540
DB 481 KWTIMSTDDKATYVMAKOSFNPTEAEAPLVNLMGSEFIDVRFQNFIELGTEGARY 540
QY 541 EKRFVWAGISNVLSHSGRENORRFRHVSAGAVGASTRMPGCDTSLGFAQLFARDKDYF 600
DB 541 EKRFVWAGISNVLSHSGRENORRFRHVSAGAVGASTRMPGCDTSLGFAQLFARDKDYF 600
QY 601 MNTNFAKTYAGSLRLOHDSILYSVSIILGEGLEIRELLPYVSKTLPCEFYGOLSTGHD 660
DB 601 MNTNFAKTYAGSLRLOHDSILYSVSIILGEGLEIRELLPYVSKTLPCEFYGOLSTGHD 660
QY 661 HRMKTESLPPPP 672
DB 661 HRMKTESLPPPP 672

```

```

RESULT 3
H86546
polymorphic outer membrane protein G family [imported] - Chlamydia pneumoniae (st
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: H86546
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: H86546
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-928 <STO>
A:Cross-references: GB:BA000008; NID:98978822; PIDN:BA98658.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: pmp_11

```

```

Query Match 33.3%; Score 1429.5; DB 2; Length 928;
Best Local Similarity 35.8%; Pred. No. 1e-74;
Matches 339; Conservative 147; Mismatches 336; Indels 125; Gaps 15;

```

```

QY 1 MKIPRFLISLVPTLMSNLGAATTEELASNSFDGTTSTTSKTSATDGTNYF 60
DB 1 MKTSLPWLVSVAIFS-CHLOSLANEELISPDSDFNCGNIDGFTTPKTS- ---TYSL 55
QY 61 KDSVVIENPKTGEGOSTSCFKNDAAGDLNLFGGGFSFTFENIDATTAAGAAIGSEAN 120
DB 56 TGDVEFFE- PGKGTPLSDSCFK- --TTDNLTFLGNGHSLTFEFDGTHAGAA-ASTTAN 111

```



**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

QY	11	SLVPTLMSNLGLAAATTEELSAANSNPDGTTSTSTSPFSKTSNADGKNVFXKSVYLE----	67
Db	466	SIMFEXSNMAY--ALFNEFDISVSG-----GSNFEFLNASSSNIQT-----PGVITIKQN	513
QY	68	-----NVPKTGETQSTSCFKNDAAAGDLNFLGCGFSFFESNIDATTA---SGAAIG	115
Db	514	FNWGSSTLMLKAGSSTETAFSIENDL---NLNATGQ--NITRQVEGDSRVKNGV--	565
QY	116	SEANKRYVTLGSFALSSELKSPASYTNGLGATNVKGNLSL-----LNDKYL-----I	164
Db	566	--AAKKNITRKQ--GNTEFGQAKATTEIKGNVITKNFTNATLKGANFAEKSPDLNAGNVI	622
QY	165	QDNSTDDGAGNACAGSLKIANNKSLSTFI-----GNSSTRGCA-----	203
Db	623	NNGNLUTTAGSILINAGNLVYSKCANQAITNTFNWAGSFDDNNGANSNISIARGAKAFDI	682
QY	204	IHKNLTLSSGGEFLFOGNTAPPAAGKGAIALADSGTSLISGSDGDIIFBQNTIGATGT	263
Db	683	NNVTSLLNTTNSDTYRTIKNISKMSQDLNITDKS-----DAEIOGINSIQKEGN	736
QY	264	VSHSAIDLGSAKITLALRAAOGHTIFYDPIVGTGTSVADALNINSPDGTQDNKEYYGTI	323
Db	737	LTISSEKVNITNQTITKAGVEG-----GRSDSEHANNLITQTEKELKLAGDL	784
QY	324	VFSGEKLTAEANDKERNKT--SKLLONVAFKNGVYLVKAGDVLNANGFSDOANSKLLMDL	381
Db	785	NISGFNAEITAKNKSODLTIGNASGDNADAKVYTPDKVXDSITSDGNVTLNSEPKVN	844
QY	382	GTSLVANTESELT-----NLEINIDSLNRGKKIKLSAA---TAQKDIRIDRPVYLAID	433
Db	845	GSSNAGNDNSTGLTISAKDVTVN--NNVTSHKTINISAAAGNVTTEGTTIN---ATTGCV	900
QY	434	ESFYONGFLMEDHSYGOIL-----ELDACKDVIYASDSRSIAVAGSPYGOGKMTIMNST	488
Db	901	EYTRAKNTIGNTISQNVYVTAJENLVTEEMAVINAKTSQVN-----ISKRT	947
QY	489	DDKRAVYSAKOSFNPTAEQAEAPLVNLLMGSFIDVRPQNFIELGTGAPYERKRFVWAG	548
Db	948	GDIGKIESTSGVNNVTASGNTLKVSNIT--GQDVVYTAAGALTT--TAGSTISA---TTG	1002
QY	549	ISNVLHSGRENQKFRHRYSGCAVYVASTRMGCGDTLSIG	588
Db	1003	NANITTKTGIDINGV--ESSSSGIVTLVAI-----GATLAAG	1036

```
Search completed: October 2, 2001, 03:27:47
Job time: 7229 sec
```

```

: Sequence 2, Application US/08719641
: Patent No. 6218141
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESS: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/719,641
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-625
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1536 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE:
: US-08-719-641-2

```

Query Match 4.8%; Score 205; DB 4; Length 1536;  
 Best Local Similarity 23.3%; Pred. No. 2.3e-08;  
 Matches 118; Conservative 74; Mismatches 185; Indels 130; Gaps 23;

```

QY 10 ISLVPTLSNMLGATTEE-----LSASNSFDGTTT-TTSES-----SKTSS----- 51
DB 1037 LKLTODLINSFNKAKETAKKDSGLTIGNTSADGTNAKVTENQKDKISADGHKVTL 1096
QY 52 ----ATDGINVYFKS-----VIEVNPRTGETQSTISCRNDAAGDLNF-LGGGGS 98
DB 1097 HSKVETSGSNNTTDESSDNAGLTLDKAVTNMNTTSHKAVSISATSGEITTKGTITN 1156
QY 99 FTFNSNDATTAAGAAIGS-EAANKTVTLT---GFSALSFELKSPASVYVINGLAINVKGL 154
DB 1157 ATTGNVEITAGTGSILIGLESSGSVTLTATGALAVNSISGNTVYVYVANSGLTTLAOS 1216
QY 155 SLDDNDKVLIDNFSTGDGAINCAGSLKIANKSLISFGNS-----SSTR 200
DB 1217 TIKGTESVTTSS--GSDIGGTISGTVTEVKATESLITQSNKIKATGTSEAVNTSATGTI 1274
QY 201 GGAI--HTNLTLSGSGELLFQGNAPTAAGKGAIAIDSGTSLISGSGGIIIEGNTI 258
DB 1275 GGITSGNTVNTANAGDLTV--GNGAEINATEGAATLTTSKGLTTEASS-----HIT 1325

```

```

QY 259 GATGVSHSAIDIGTSAKITLRAAGHTIYFDPITVGTSTVADALNINSPDTGDNKE 318
DB 1336 SAKGVNLSAOGSVAGSINAAVNT-----LNTGTITLYKGSINNT----- 1368
QY 319 YGTITVSGEKLTEAEADKERNRTSKLLQNVAFKNGTVYVLLKGDVYLSANGFSODANSKLI 378
DB 1369 -SGTVIVN-----AKDAE-----LNGAALGNHTVY-----NATNANG----- 1399
QY 379 MGLGTSIVANTES-IETNLEINIDSL---RNGKKIKTSATACKDRIDRPVYIAISD 433
DB 1400 ---SGSVIATTSRVNITGDLTTLTGLNITISKNGINTVYV-----LKGVI----- 1442
QY 434 ESFYONGFLNEDHSYDGLIELDAGDI 460
DB 1443 VKYIQGIASVDVEIARILKVKDL 1469

```

```

RESULT 15
US-08-728-470-9
: Sequence 9, Application US/08728470
: Patent No. 5928651
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESS: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/728,470
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-633
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1338 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-728-470-9

```

Query Match 4.7%; Score 200.5; DB 2; Length 1338;  
 Best Local Similarity 22.0%; Pred. No. 4.5e-08;  
 Matches 141; Conservative 99; Mismatches 269; Indels 131; Gaps 27;

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:  
US-08-728-470-2

Query Match 4.8%; Score 205; DB 2; Length 1536;  
Best Local Similarity 23.3%; Pred. No. 2.3e-08;  
Matches 118; Conservative 74; Mismatches 185; Indels 130; Gaps 23;

QY 10 ISLVPTLSMNLGAATTEE-----LSASNSPDGTTT-TTSFS-----SKTSS----- 51  
DB 1037 LKLVQDINISGFNKAETAKGSDLTIGNTNSADGTNAKKVTFNQVDSKISADGHKVTL 1096  
QY 52 ----ATDGTNVFKDS-----VVIENVPKGTGETOSTSCFKNDAAAGLNF-LGGGFS 98  
DB 1097 HSKVETSGSNNTEDSSDNNAGLTIDAKNVTVNNNITSHKAVSISATSGEITTKGTGTIN 1156  
QY 99 FTFSNIDATTAAGAAIGS-BAANKTVTLTSGSALSPKSPASTVTNGIGAINVGNL 154  
DB 1157 ATGNVETITAGTISGVTILGIESSSGVTLTATGALAVSNISGNTVTVYVANSALTLTLAGS 1216  
QY 155 SLIDNDKVLQDNESTGDGAINCAGSLKIANKSLSFIGNS-----SSTR 200  
DB 1217 TIKTESVTSS--QSGDIGTISGTVKATESLTQSNKIKATVTEANVTSAATGTI 1274  
QY 201 GGAL--HTKNULTSSGGETLFOGNTAPTAAGKGAIAIADSGTISGDSGDIIFEGNTI 258  
DB 1275 GGTTSGTAVNTAAGDLTV--GNGAEINATEGATLTTSCKLITTEASS-----HIT 1325  
QY 259 GATGTVSHSAIDLGTSAKITALRAAQHTIYFDPIVTGTSVADALINSPDTGDNKE 318  
DB 1326 SAKQVNLMSADGVSAGSINAAVNT-----LMTGTGLTVKGSINAT----- 1368  
QY 319 YTGTVESGKLEBAEKDEKNTSKLLQNVAFKNGTVVLKGDVYLSANGFSODANSKLI 378  
DB 1369 -SGTLVIN-----AKDAE-----LNGAALGNHTVV-----NATVANG----- 1399  
QY 379 MDLGTSLVANTES-IELTNEINISL-----RNGKKIKLSAATAOKDIRDRPVLAISD 433  
DB 1400 ---SGVIATTSKRVNITGDLITINGINITSKGINTVL-----LKGVKI-----D 1442  
QY 434 ESFYONGFLNEDHSYDGLLELDAGKDI 460  
DB 1443 VKYIQPGIASVDEYIEAKRILEKVKDL 1469

RESULT 13  
US-08-617-697-2  
Sequence 2, Application US/08617697  
Patent No. 5977336  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/617,697  
FILING DATE: 01-Apr-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 05-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-557  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-617-697-2

Query Match 4.8%; Score 205; DB 2; Length 1536;  
Best Local Similarity 23.3%; Pred. No. 2.3e-08;  
Matches 118; Conservative 74; Mismatches 185; Indels 130; Gaps 23;

QY 10 ISLVPTLSMNLGAATTEE-----LSASNSPDGTTT-TTSFS-----SKTSS----- 51  
DB 1037 LKLVQDINISGFNKAETAKGSDLTIGNTNSADGTNAKKVTFNQVDSKISADGHKVTL 1096  
QY 52 ----ATDGTNVFKDS-----VVIENVPKGTGETOSTSCFKNDAAAGLNF-LGGGFS 98  
DB 1097 HSKVETSGSNNTEDSSDNNAGLTIDAKNVTVNNNITSHKAVSISATSGEITTKGTGTIN 1156  
QY 99 FTFSNIDATTAAGAAIGS-BAANKTVTLTSGSALSPKSPASTVTNGIGAINVGNL 154  
DB 1097 HSKVETSGSNNTEDSSDNNAGLTIDAKNVTVNNNITSHKAVSISATSGEITTKGTGTIN 1156  
QY 155 SLIDNDKVLQDNESTGDGAINCAGSLKIANKSLSFIGNS-----SSTR 200  
DB 1217 TIKTESVTSS--QSGDIGTISGTVKATESLTQSNKIKATVTEANVTSAATGTI 1274  
QY 201 GGAL--HTKNULTSSGGETLFOGNTAPTAAGKGAIAIADSGTISGDSGDIIFEGNTI 258  
DB 1275 GGTTSGTAVNTAAGDLTV--GNGAEINATEGATLTTSCKLITTEASS-----HIT 1325  
QY 259 GATGTVSHSAIDLGTSAKITALRAAQHTIYFDPIVTGTSVADALINSPDTGDNKE 318  
DB 1326 SAKQVNLMSADGVSAGSINAAVNT-----LMTGTGLTVKGSINAT----- 1368  
QY 319 YTGTVESGKLEBAEKDEKNTSKLLQNVAFKNGTVVLKGDVYLSANGFSODANSKLI 378  
DB 1369 -SGTLVIN-----AKDAE-----LNGAALGNHTVV-----NATVANG----- 1399  
QY 379 MDLGTSLVANTES-IELTNEINISL-----RNGKKIKLSAATAOKDIRDRPVLAISD 433  
DB 1400 ---SGVIATTSKRVNITGDLITINGINITSKGINTVL-----LKGVKI-----D 1442  
QY 434 ESFYONGFLNEDHSYDGLLELDAGKDI 460  
DB 1443 VKYIQPGIASVDEYIEAKRILEKVKDL 1469

RESULT 14  
US-08-719-641-2





```

? COUNTRY: U.S.A.
? ZIP: 22202-0286
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/302,832
? FILING DATE: 16-SEP-1994
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: GB 9205704.1
? FILING DATE: 16-MAR-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US pct/us93/02166
? FILING DATE: 16-MAR-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Berkstresser, Jerry M
? REGISTRATION NUMBER: 22,651
? REFERENCE/DOCKET NUMBER: 1038-404
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 415-0810
? TELEFAX: (703) 415-0813
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1536 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-08-302-832-2

```

```

Query Match 4.8%; Score 205; DB 1; Length 1536;
Best Local Similarity 23.3%; Pred. No. 2.3e-08;
Matches 118; Conservative 74; Mismatches 185; Indels 130; Gaps 23;

QY 10 ISLVPTLSMNLGAATTEE-----LSASNSFDGTS--TTSFS-----SKTSS----- 51
DB 1037 LKLTODLNISGPNKAEITADGSDLTIGNTNSADGTVNNAKVTFNQVYKDSKISADGKRVTL 1096
QY 52 ----ATDGTNYVFKDS-----VVIENVPKGTGETOSTCFKNDAAAGLNF-LGGGFS 98
DB 1097 HSKVETSGSNNTEDSDNNAGLTIIDAKNVTYNNNITSHKAVSISATSGEITTTGTGTTIN 1156
QY 99 FTFSNIDATTAAGAIGS-EAANKVTLS---GFSALFLKSPASTVTNGLAIVKGNL 154
DB 1157 ATGCVNVEITATQTSIIIGIIESSGVSITLFTATEGALAVSNISGNTVTVTANSALTLTLAGS 1216
QY 155 SLDDNDKVLIDNFTSGDGAINCAGSLKIANKKSIFGNS-----SSTR 200
DB 1217 TIKGTESVTTSS--QSDDIGTISGTVVEKATSELTQSNKIKATTTGANTYSATGTI 1274
QY 201 GGA1--HTKMLTSSGGETLFOGNTAPTAAKGGAIAIDSGTSLISGDSGDIIFEGNTI 258
DB 1275 GGTISGNTVAVTANAGDLTY--GNGAFINATEGAATLTSSGKLTITRASS-----HIT 1325
QY 259 GATGTVSHSAIDIGTSAKITALLRAAGHTIIFYDPTVTGTSVADALINSPDGDNKE 318
DB 1326 SAKGOVVLISADQDSVAGSINAAVNT-----LNTTGTLTFTVKGSNINAT----- 1368
QY 319 YTGTVSSGKLTLEAKKDEKNTSKILLONAVAFKNGTVYVYKGDVYVLSANFSODANSKLI 378
DB 1369 -SGTLVYN-----AKDAE-----LNGAALGNHTV-----NATNANG----- 1399
QY 379 MDGTSIVANTES-IETNLEINISL-----RNGKIKILSAATAOKDIRIDRPVLAISD 433
DB 1400 ---SGVIAATSSRVNATGDLITINGNITSKGINVY-----LKGVKI-----D 1442
QY 434 ESFYONGFLNEDSHSYDGLIELDAGKDI 460
DB 1443 VKYIOPGIASVDVEIEKRIIEKVKDIL 1469

```

```

RESULT 10
US-08-530-198-2
? Sequence 2, Application US/08530198
? Patent No. 5869065
? GENERAL INFORMATION:
? APPLICANT: BARENKAMP, STEPHEN J
? APPLICANT: ST. GEME III, JOSEPH M
? TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
? TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Shoemaker and Mattare, Ltd
? STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
? STREET: Bldg. 1
? CITY: Arlington
? STATE: Virginia
? COUNTRY: U.S.A.
? ZIP: 22202-0286
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/530,198
? FILING DATE: 13-DEC-1995
? CLASSIFICATION: 424
? ATTORNEY/AGENT INFORMATION:
? NAME: BERKSTRESSER, JERRY M
? REGISTRATION NUMBER: 22,651
? REFERENCE/DOCKET NUMBER: JWB-1186
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 415-0810
? TELEFAX: (703) 415-0813
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1536 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-530-198-2

Query Match 4.8%; Score 205; DB 2; Length 1536;
Best Local Similarity 23.3%; Pred. No. 2.3e-08;
Matches 118; Conservative 74; Mismatches 185; Indels 130; Gaps 23;

QY 10 ISLVPTLSMNLGAATTEE-----LSASNSFDGTS--TTSFS-----SKTSS----- 51
DB 1037 LKLTODLNISGPNKAEITADGSDLTIGNTNSADGTVNNAKVTFNQVYKDSKISADGKRVTL 1096
QY 52 ----ATDGTNYVFKDS-----VVIENVPKGTGETOSTCFKNDAAAGLNF-LGGGFS 98
DB 1097 HSKVETSGSNNTEDSDNNAGLTIIDAKNVTYNNNITSHKAVSISATSGEITTTGTGTTIN 1156
QY 99 FTFSNIDATTAAGAIGS-EAANKVTLS---GFSALFLKSPASTVTNGLAIVKGNL 154
DB 1157 ATGCVNVEITATQTSIIIGIIESSGVSITLFTATEGALAVSNISGNTVTVTANSALTLTLAGS 1216
QY 155 SLDDNDKVLIDNFTSGDGAINCAGSLKIANKKSIFGNS-----SSTR 200
DB 1217 TIKGTESVTTSS--QSDDIGTISGTVVEKATSELTQSNKIKATTTGANTYSATGTI 1274
QY 201 GGA1--HTKMLTSSGGETLFOGNTAPTAAKGGAIAIDSGTSLISGDSGDIIFEGNTI 258
DB 1275 GGTISGNTVAVTANAGDLTY--GNGAFINATEGAATLTSSGKLTITRASS-----HIT 1325
QY 259 GATGTVSHSAIDIGTSAKITALLRAAGHTIIFYDPTVTGTSVADALINSPDGDNKE 318
DB 1326 SAKGOVVLISADQDSVAGSINAAVNT-----LNTTGTLTFTVKGSNINAT----- 1368

```

```

QY 19 SNLGAATTEELASNSPDGTTSTSESSKTSANDGTNYVFKDSVIEVNPKTGTOST 78
DB 684 SSKGLTTOYRSSAGVNGVNGMNSFNK-----EGAVYNRK-----LKPENMNT 730
QY 79 S-----CFKNDAAGD-----LNFLLGGGFFSFFSNIDATTAAGAIQSEA----- 118
DB 731 SRPLPIRPLANTATGCGSVFPDIYANHSRGAEKMKSEINISNGANFLNHSVRDDAF 790
QY 119 -ANKVTLSGFSALFLKSPASTVTNGL--GAINVKGNISLNDKVLIDNEFTDGA 175
DB 791 KINKDLITINATNSFSLROTDFDYVARNAINSTYNNISIL-----CGN 835
QY 176 INCAGSLKANKKSLFISNSTRGCAIHTKNTLLSSGGEILFOGNTAPTAAG----- 229
DB 836 VTLGGQ-----NSSSSITG-----NITTEKANTVLEANNAPNOONIRDRYI 877
QY 230 KCGATAIADSGTSLISGDSGDI-----IFEG--NTIGATG-----TVSH 266
DB 878 KLGSLV--NGSLSLTGENADIKGNLTISESATPFKGTNDLITNGFTNNGTAELNITQ 935
QY 267 SAIDGT-----SAKITLARAAGHTIYDPITYTGTSVADALNINSPDTGDN--KEYT 320
DB 936 GYVKGIGNTNDGLNITTTAKRNQRSIIIGDIIKKGSINTIDSNDAEIOIGNISQKE 995
QY 321 GTIVFSGEK-----LLEAKDEKRTSKLONVAFKNGTVLKGDVLSANGF----- 369
DB 996 GNLTISDKINTIKQTIKKGIDGEDSSDATSANMLTKTEKLTEDLSISGFKAEI 1055
QY 370 -----SODANSKLINDLGTSLVANTESELINL--EINID--SLRNGKIKL 412
DB 1056 TAKDGRDLTGNNSDNGSC-----AEAKTVTFNNVVKSKISADGHNTVLSNKVKT 1105
QY 413 SAATAKDIRIDRPVVLASDESFYONGFLNEHSDGLELIDAGDIYISADSRINAV 472
DB 1106 SSSNGGRESNDGTLTIT--AKNVEVNRDITSLKTVNITASKAKVTTAGS--TINAT 1160
QY 473 QSPYGVQWNTIMSTDDKATVSNAKOSEFNPTAEQEAFLVPLNLGSPFLDVRPFQNFIE 532
DB 1161 -----NGRASITTKTGDISGITISGNTVSVAT-----VD 1189
QY 533 LCTEGAPYKRFVWAGISVNLHRSRENORRFRHVSAGAVGASTBMP--GGDTLSLG 588
DB 1190 LTTKSG-----SKTEAKSGEANTVTSATGTIGTISGNTVAVTANAGD--LTVG 1235

```

RESULT 8  
US-08-038-682-2  
Sequence 2, Application US/08038682  
Patent No. 5549897

## GENERAL INFORMATION:

APPLICANT: BARENKAMP, STEPHEN J  
APPLICANT: ST. GEME III, JOSEPH W  
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Shoemaker and Mattare, Ltd  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/038,682  
FILING DATE: 16-MAR-1993  
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: BERTSTRESSER, JERRY W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-293  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-038-682-2

Query Match 4.8%; Score 205; DB 1; Length 1536;  
Best Local Similarity 23.3%; Pred. No. 2,3e-08;  
Matches 118; Conservative 74; Mismatches 185; Indels 130; Gaps 23;

```

QY 10 ISLVPTLSNMLGAATTE-----LSANSFPGTTS--TTSS-----SKTS----- 51
DB 1037 LKITDPLNISGFENKKEITAKDSDLTIGTNTSADGTNAKVFENQYKDSADGHRVTL 1096
QY 52 -----ADGTNYVFKDS-----VYIENVPKGTGTOSTSCFNDAAGDLNF--LGGS 98
DB 1097 HSKVETSGSNNTTEDSDNNAGLITIDAKNVTYNNNTTSKRAVISATSGEITTKGTITT 1156
QY 99 FTFSNIDATTAAGAAIGS--EAANKVTVLS--GFSALFLKSPASTVTNGLAINVKNL 154
DB 1157 AVTGNVEIQAQSGISLIGSSSGSVTLTATGALAVNISGWTVAVTANSGALTLTLAGS 1216
QY 155 SLINDKVLIDNFTSGDGAINCAGSLKIANKKSIFGNS-----SSTR 200
DB 1217 TIKGSEVYTTSS--GSGDIGGTISGTVKATESLTTQSNKIKATTEGANTVTSATGTI 1274
QY 201 GGAI--HTKNLTLSGGEILFOGNTAPTAAGKGAIAIDSGTSLISGSDIIFEGNTI 258
DB 1275 GGTISGNTVAVNAGDLTV--GNGAEINATECAATLTSSGKITTEAS-----HIT 1325
QY 259 GATGVVSHSAIDLGTSAKITLARAAGHTIYFDPITVTGTSVADALNINSPDTGDNKE 318
DB 1326 SAKGVVNLISAQGSVAGSINAANVT-----LNTGTTLTVGNSINAT----- 1368
QY 319 YGTIVFSGEKLTEAEKKEKRTSKLONVAFKNGTVLKGDVLSANGFQDANSKLI 378
DB 1369 -SGTLVIN-----AKDAE-----LNGAALGNHTTV--MATNANG----- 1399
QY 379 MDLGTSLVANTES--IELTMEINIDSL--RNGKKIKLSAATAOKDIRIDRPVVLASD 433
DB 1400 ---SGSVATITSSRVITDGLTINGLNIIISKNGINTVL-----LKGVAI-----D 1442
QY 434 ESFYONGFLNEHSDYDGLELIDAGKDI 460
DB 1443 VKYIOPGIASVDEVIEAKRILEKVDL 1469

```

## RESULT 9

US-08-302-832-2  
Sequence 2, Application US/08302832  
Patent No. 5603938

## GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia

```

APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
  ADDRESSEE: Shoemaker and Matlare, Ltd.
  STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
  CITY: Arlington
  STATE: Virginia
  COUNTRY: U.S.A.
  ZIP: 22202-0286
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  OPERATING SYSTEM: IBM PC compatible
  SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/617,697
  FILING DATE: 01-APR-1996
  CLASSIFICATION: 424
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/302,832
    FILING DATE: 05-OCT-1994
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US PCT/US93/02166
      FILING DATE: 16-MAR-1993
      ATTORNEY/AGENT INFORMATION:
        NAME: Berkstesser, Jerry W
        REGISTRATION NUMBER: 22,651
        REFERENCE/DOCKET NUMBER: 1038-557
        TELECOMMUNICATION INFORMATION:
          TELEPHONE: (703) 415-0810
          TELEFAX: (703) 415-0813
          INFORMATION FOR SEQ ID NO: 4:
            SEQUENCE CHARACTERISTICS:
              LENGTH: 1477 amino acids
              TYPE: amino acid
              STRANDEDNESS: single
              TOPOLOGY: linear
US-08-617-697-4

Query Match
Best Local Similarity 20.7%; Score 205; DB 2; Length 1477;
Matches 136; Conservative 88; Mismatches 240; Indels 194; Gaps 28;

```

```

Db 996 GNLTISSDKINITKQITIKKIGDESSDPATSNANLTIKTELKTEDLSISGFNKAELI 1055
Qy 370 -----SODANSKLMIDGLTSLVANTESELNLN---EINID--SLRNGRKIKL 412
Db 1056 TAKDGRDLTIGNSNDNGSG-----AEAKVTNNVNDKSLISDGHVTLNSKVKT 1105
Qy 413 SAATAKDIRIDRPVVAISDESFQNGFLNEDHSYDGLLELDAGKDIVISADRSINAV 472
Db 1106 SSSNGGRESNDNDOTGLTIT---AKNVEVNDKDTLSLKTIVNITASEKVTYTAGS-TINAT 1160
Qy 473 QSPYGOQKWTINWSTDCKATVSWAKOSFNPFAEQAPLVPNLLGSPFDVDPFONFIE 532
Db 1161 -----NGKASITTKTGDISGTTISGNTVSVSAT-----VD 1189

Qy 533 IGTGAPYERKRFVWAGISNVLRSGRENOKRFRVSGAVVAGSTAMP--GGDTLSIG 588
Db 1190 LTTKSG-----SKIEANSGEANVTSACTIGCTTISGNTVYANNGD-LTVG 1235

RESULT 7
US-08-719-641-4
; Sequence 4, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matlare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstesser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-719-641-4

Query Match
Best Local Similarity 20.7%; Score 205; DB 4; Length 1477;
Matches 136; Conservative 88; Mismatches 240; Indels 194; Gaps 28;

```

QY	79	S-----CFKNDAAD-----	LNFGGSEFSPFNSIDATTSAGAISEA-----	118	
Db	731	SKULPIRFILNATITATGGGVFYDIYANHSGRGAEIKMEISINSGANFTLLSHVRGDDAF		790	
QY	119	-ANKVTYLSGFSALSPKSPASTVNTGL--GAINVKGNLSLIDNDKVLIDONFSTGDGA	175		
Db	791	KINKDLIINNTNSNFSLRQTKDFYDGARNAINSTYNSIL-----	GGN	835	
QY	176	INCAGSLIKIANKSLSPFNGSSSTRGCAIHYTNLTLSSGGETLFOGNAPTPAG-----	229		
Db	836	VTLGGQ-----	NSSSITG-----NITIEKANVTLEANNAPMOONIRDVI	877	
QY	230	KGGAIAIDSCITLISGDSGDI-----	IFPS-----NITGANG-----	TYSH	266
Db	878	KLSGLV--NGSISLTGENADIKGNLTISESATPFKGRDRTLINTIGNFTNGTAINTQ	935		
QY	267	SAIDLCF-----SAKTALRAAQHTIYEDPLTVYGTSPVADALINSPPTGDN--KEYT	320		
Db	936	GVKLGNVTGDDGLNITTHAKRNQBSIIIGGLIINKKSLINIDSNNDAEIQIGAISSKE	995		
QY	321	GTIVFSGE-----LTEAEADENKRSIKLQNAFKNGVYLVKGDVYLSANGE----	369		
Db	996	GNLTISDKINTIKQJTIKKIGIDFESSDSDATSNALFTKYTELKTLDLISGPNKAEI	10555		
QY	370	-----SODANSKLINDLGLSVLANTESTELTNL--	ELNID--SLRNCKAKKL	412	
Db	1056	TAKDGRDLTIGNSDNSG-----	AEAKTYFNNVNDKSLAAGHNVTLSMKRYKT	1105	
QY	413	SAATAOKDIRIDRPVVALISDESFYNGFLANDHSDGLIEDAKDIYISADRSINAV	472		
Db	1106	SSSNGGESHSDMDJGLTIT-----	AKNEVNMJDITSIKTYNTTAAEKYTTAGS--TIAT	1160	
QY	473	QSPGYOGKWTIMWSTDDKKAIVSWAKOSFNPJAOEAPLVNLLMGSFIDVRPQNFIE	532		
Db	1161	-----NGKASITTKTGDISGITSGTIVSVSAT-----	VD	1189	
QY	533	LGTEGAPYEREFVVAISIVNLAHRSNGENRKRHHVSGGAVVCASIRMP--	GGDTLSLG	588	
Db	1190	LTTKSG-----	SKTEAKSGGANVSATGTITGIGISNTVYVANNAGD--LTVG	1235	

RESULT 5  
US-08-728-470-4  
; Sequence 4, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728.470  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993

```

? PROR APPLICATION DATA:
? APPLICATION NUMBER: GB 9205704.1
? FILING DATE: 16-MAR-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Belkstresser, Jerry W
? REGISTRATION NUMBER: 22,651
? REFERENCE/DOCKET NUMBER: 1038-633
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 415-0810
? TELEFAX: (703) 415-0813
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1477 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
? US-08-728-470-4

```

```

Query Match      4.8%: Score 205; DB 2; Length 1477;
Best Local Similarity 20.7%: Pred. No. 2.2e-08;
Matches 136; Conservative 88; Mismatches 240; Indels 194; Gaps 28;

OY 19 SNLLGAATTEELASNSPDGTTSTSPFSKSSAIDGTNYVKDSVLENYPKTEGTOST 78
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 684 SNSKGLITQYSSAGVNVNGVGNNSFNLK-----EGAKVNRK-----LKPENMMT 730
OY 79 S-----CFKDDAAGD-----LNFLLGGSEFTFSNIDATVSGAIGSEA----- 118
DB 731 SKPLPIRELAITATGGSVFEDDIYANHSGRGAELKMEKSEINISNAGNPLNHSVGGDDAF 790
OY 119 -ANTVTLISGRSALSFLKSPASTYNGL--GAINKGNLSLIDNDKVLQIDNFSRGDGA 175
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 791 KINDDLTINATNNSFSLKQTKODEFYDGAARNINSTYINISL-----GGN 835
OY 176 INCAGSLKIANNNLSLFIKSSSTRGCAIHTNMLLSGGELLFOGNAPTAA----- 229
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 836 VTLGQ-----NSSSITG-----NIIERAAVNTLEANNPNOONRDVY 877
OY 230 KGAIAIADSGTSLISGDSGI-----IFEG--NTIGATG-----TVSH 266
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 878 KLGSLV--NGLSLITGENMADIKGNLTJESATFPKGRDPLNTIGNFTNNGAEINITQ 935
OY 267 SAIDLGT-----SAKITPLRAAQGHTIIFYDEITYTGSIVADALINSPTGDN--KEYT 320
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 936 GVVKLGNVNTNDGLNLTTHAKRNORSIIIGDIINKKGSINTIDSNDAEIOIGGNISQKE 995
OY 321 GTIYFSGEK-----LFEAEAKDEKRNPTSKLQNVAFKNGVYVLKGDVYLSANGF----- 369
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 996 GNLTISSDKINITQOITIKKGIDGEDSSDATSNANLTIKTELTLTEDLSISGFKAEI 1055
OY 370 -----SODANSKLIMDLGTSLVAMTESIELTNL--EINID--SLRNKGKIKL 412
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1056 TAKGRDLITIGNSNDGNSG-----AEKATYFNNAVDSKISADCHNVTNLSKYVT 1105
OY 413 SAATAQKDRIIDRPVYLAISDESFYQNGFLEHDSYDGLLELDACKDIYISADSINAV 472
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1106 SSSNGGRSNDNDPGLGIT--AKNEVYNNKDITSLKTVNITASEKVTYTAGS--TINT 1160
OY 473 QSPFYGCKWTIINMSTDKKATVSMAKOSFNPTLAQDEAPLVNLLMGSITDVRPFONFIE 532
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1161 -----NKAASITTKTGIGISGITSGNTSVSAT-----VD 1189
OY 533 LCTEGAPYEKFPWVAGISENVLHRSGREVQRKFRHVSAGVAGASTRMP--GGDTLSLG 588
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1190 LTTXSG-----SKTEAKSGEANVTSAATGTTIGTISGNTVNVTAACD--LTVG 1235

RESULT 6
US-08-617-697-4
Sequence 4, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:

```

```

? APPLICANT: ST. GENE III, JOSEPH W
? TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: Shoemaker and Mattare, Ltd
? STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
? CITY: Arlington
? STATE: Virginia
? COUNTRY: U.S.A.
? ZIP: 22202-0286
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/530,198
? FILING DATE: 13-DEC-1995
? CLASSIFICATION: 424
? ATTORNEY/AGENT INFORMATION:
? NAME: BERTSTRESSER, JERRY W
? REGISTRATION NUMBER: 22,651
? REFERENCE/DOCKET NUMBER: JMB-1186
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 415-0810
? TELEFAX: (703) 415-0813
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1477 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-530-198-4

```

Query Match 4.8%; Score 205; DB 2; Length 1477;

Best Local Similarity 20.7%; Pred. No. 2.2e-08; Matches 136; Conservative 88; Mismatches 240; Indels 194; Gaps 28;

```

QY 19 SNLGAATTELSASNSFDDTSTTSSTSSATDGTNVFKDSVYIENVPKGTGTOST 78
DB 684 SNSGGLTTOYRSSAGVFNVCNGMSFNK-----EGAKVNFK-----LKNENMNT 730
QY 79 S-----CFRNDAAAGD-----LNFLLGGSEFTFSNIDATTAAGAAIGSEA----- 118
DB 731 SKPLPIRFPLANITATGGSVFEDIYANHSGRGAEIKMSEINISNGANFTLNSHVRGDAF 790
QY 119 -AKTVTLSGFSALSFKSPASTYTNGL--GATVWKGNSLIDNDKYLIDNESTGCGGA 175
DB 791 KINKDLINATNSNFSLRQRDEYDYARNAJNSTYNISIL-----GGN 835
QY 176 INCAGSLKIANNSLSLFSRGGAIHTKNTLSSGGETLFOGNVAPTAAAG----- 229
DB 836 VTLLGGQ-----NSSSSITG-----NTTIKKAANTVTEANNAPOQNIIRPRI 877
QY 230 KGAALADSGTSLISGSDI-----IFEG--NTIGATG-----TVSH 266
DB 878 KLSLAV--NGSLSLTGENDIKGNLTISSEAFPKGTRDPLNITNGFTNGTAEINITO 935
QY 267 SAIDLGT-----SAKITALAAAGHTIFYFDPTVTVSTSVADALNTNSPTGDN--KEYT 320
DB 936 GYVKGAVNTNDGLNITTHAKRNQORSIIIGDDIINKKGLNITSSNNAETIOGIGNISOKE 995
QY 321 GTYVSEK-----LTEADAKDEKNTSKLQVAFKNGTVVVGKGVLSANGF----- 369
DB 996 GNLTISSDKINIKQIIRKIGIDGEDSSDATSNANLITKTELKLELIDSLISFPNNAEI 1055
QY 370 -----SODANSKLIMDLGTVSLVANTSEIELTNT--EINID--SLRNGKRIKL 412
DB 1056 TAKDGRDLTIGNSDGNSG-----AEAKTVTFENNVKDSKISADGHNVTLSNIVKT 1105

```

```

QY 413 SAATAQDIRIDRPVLAISDESEFYONGFLNEDHSYDGLLELAGKDIVISADRSINAV 472
DB 1106 SSSNGGREGNSDNDGTGLTIT-----AKNVEVKKDITLSKTVITASEKVTYTAGS--TINAT 1160
QY 473 OSPYOGKWTINWSTDDKATVSMAKOSEPNPTAEQEAFLVPLNMGSEFIDRPPQNFIE 532
DB 1161 -----NKGASITTKTGDISGTISGNTVSVAT-----VD 1189
QY 533 LGTEGAPYERKFWAGLSNVLAHRSRGRENQKFRHVSAGAVVASTRMP--GGDTLSIG 588
DB 1190 LTTKSG-----SKIEAKSGEANYTSATGTIGTIGTNTVNTANAGD--LTVG 1235

```

#### RESULT 4

```

? US-08-469-880-4
? Sequence 4, Application US/08469880
? Patent No. 5876733
? GENERAL INFORMATION:
? APPLICANT: Barenkamp, Stephen J.
? TITLE OF INVENTION: High Molecular Weight Surface Proteins
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: Shoemaker and Mattare, Ltd.
? STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
? CITY: Arlington
? STATE: Virginia
? COUNTRY: U.S.A.
? ZIP: 22202-0286
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/469,880
? FILING DATE: 06-JUN-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: GB 9205704.1
? FILING DATE: 16-MAR-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US PCT/US93/02166
? FILING DATE: 16-MAR-1993
? APPLICATION DATA:
? APPLICATION NUMBER: US 08/302,832
? FILING DATE: 16-SEP-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Bertstresser, Jerry W
? REGISTRATION NUMBER: 22,651
? REFERENCE/DOCKET NUMBER: 1038-516 MIS:Vg
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 415-0810
? TELEFAX: (703) 415-0813
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1477 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? US-08-469-880-4

```

Query Match 4.8%; Score 205; DB 2; Length 1477;

Best Local Similarity 20.7%; Pred. No. 2.2e-08; Matches 136; Conservative 88; Mismatches 240; Indels 194; Gaps 28;

```

QY 19 SNLGAATTELSASNSFDDTSTTSSTSSATDGTNVFKDSVYIENVPKGTGTOST 78
DB 684 SNSGGLTTOYRSSAGVFNVCNGMSFNK-----EGAKVNFK-----LKNENMNT 730

```

```

Db 684 SNSKGLTTOYRSSAGVNFVNGVNMSPNLK-----EGAKVNRK-----LKPENMNT 730
QY 79 S-----CFKNDAAAGD-----LNFLGGGFSFESNIDATFASGAALGSEA----- 118
Db 731 SKPLPRPLANTITATGGGVFPDIYANHSGRGAELKMSFINISNGANFTLNSHVRCDDAF 790
QY 119 -ANKVTYLSGFSALFLKSPASTVTNGL--GAINVKNLSLNDKRVLLIODNESTGDGA 175
Db 791 KINKDLITNATNSFSLRQTKDDFYDGYARNALINSTYINISIL-----GDN 835
QY 176 INCAGSLKATNKKSLSFIONSSSTRGCAIHTKRLTSSGGEFLFOGNTAPTAAG----- 229
Db 836 VTLLGGQ-----NSSSSITG-----NITTEKAAVNTLEANNAPNOONIRDRVI 877
QY 230 KGAIALADSGTLISGDSGDI-----IFEG-----NTIGATG-----TVSH 266
Db 878 KIGSLIY--NGSLSLTGENADIKGNLTISESATFKKTRDTLITGNFTNNGTAELNITQ 935
QY 267 SAIDIGT-----SAKITLRAAOGHTIYEPDIYTGSTSVADALINSPDTGDN--KEYT 320
Db 936 GVVKLGNTVNDGDLNITTHAKRNORSIIIGDIIKKGSLINTDSNNDAEIQIGNISQKE 995
QY 321 GTIVSGEK-----LTEEAKDEKNRTSKLLONAFKNGTYVVKGDVYLSANGF----- 369
Db 996 GMLTISDKINITYKQITTKIGIDGEDSSSDATSNANLTITKTELKLTEDLSISGFKAEL 1055
QY 370 -----SODANSKLIMDLGTLVANTESIELTNL---EINID--SLRNGKKIKL 412
Db 1056 TAKDGRDLITGNSNDGNSG-----AEAKYTFNNVKNKSKISADCHNTLNSKVT 1105
QY 413 SAATAOKDIRIDRPVYLAISDESFYONGFLNEDHSYDGLIELDAGKDIYISDSRSINAV 472
Db 1106 SSSNGRGESNDNDGLTIT-----AKNVEVNRKDIYSLKVTNITASBKVYTTAGS--TINAT 1160
QY 473 QSPYGOQKWTINMSTDDKKAIVSNAKOSFNPTAODEALVYNLMSGTIDVRPQNFIE 532
Db 1161 -----NGKASITTKTGIDISGTLISGNTVSVAT-----VD 1189
QY 533 LCTEGAPYKRRFWVAGISNVLHRSRGRENORRFRHVSGAGVAGASTRMP--GGDTLSIG 588
Db 1190 LTTKSG-----SKIEAKSGEANTSATGTIGTISGNTVAVTANAD--LTVG 1235

RESULT 2
US-08-302-832-4
; Sequence 4, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-302-832-4

Query Match 4.8%; Score 205; DB 1; Length 1477;
Best Local Similarity 20.7%; Pred. No. 2,2e-08;
Matches 136; Conservative 88; Mismatches 240; Indels 194; Gaps 28;

QY 19 SNLGAATTEELISASNSPFGTITSTFSKTSATDGTIVYRKDSVVIENVKRTGETOST 78
Db 684 SNSKGLTTOYRSSAGVNFVNGVNMSPNLK-----EGAKVNRK-----LKPENMNT 730
QY 79 S-----CFKNDAAAGD-----LNFLGGGFSFESNIDATFASGAALGSEA----- 118
Db 731 SKPLPRPLANTITATGGGVFPDIYANHSGRGAELKMSFINISNGANFTLNSHVRCDDAF 790
QY 119 -ANKVTYLSGFSALFLKSPASTVTNGL--GAINVKNLSLNDKRVLLIODNESTGDGA 175
Db 791 KINKDLITNATNSFSLRQTKDDFYDGYARNALINSTYINISIL-----GDN 835
QY 176 INCAGSLKATNKKSLSFIONSSSTRGCAIHTKRLTSSGGEFLFOGNTAPTAAG----- 229
Db 836 VTLLGGQ-----NSSSSITG-----NITTEKAAVNTLEANNAPNOONIRDRVI 877
QY 230 KGAIALADSGTLISGDSGDI-----IFEG-----NTIGATG-----TVSH 266
Db 878 KIGSLIY--NGSLSLTGENADIKGNLTISESATFKKTRDTLITGNFTNNGTAELNITQ 935
QY 321 GTIVSGEK-----LTEEAKDEKNRTSKLLONAFKNGTYVVKGDVYLSANGF----- 369
Db 996 GMLTISDKINITYKQITTKIGIDGEDSSSDATSNANLTITKTELKLTEDLSISGFKAEL 1055
QY 370 -----SODANSKLIMDLGTLVANTESIELTNL---EINID--SLRNGKKIKL 412
Db 1056 TAKDGRDLITGNSNDGNSG-----AEAKYTFNNVKNKSKISADCHNTLNSKVT 1105
QY 413 SAATAOKDIRIDRPVYLAISDESFYONGFLNEDHSYDGLIELDAGKDIYISDSRSINAV 472
Db 1106 SSSNGRGESNDNDGLTIT-----AKNVEVNRKDIYSLKVTNITASBKVYTTAGS--TINAT 1160
QY 473 QSPYGOQKWTINMSTDDKKAIVSNAKOSFNPTAODEALVYNLMSGTIDVRPQNFIE 532
Db 1161 -----NGKASITTKTGIDISGTLISGNTVSVAT-----VD 1189
QY 533 LCTEGAPYKRRFWVAGISNVLHRSRGRENORRFRHVSGAGVAGASTRMP--GGDTLSIG 588
Db 1190 LTTKSG-----SKIEAKSGEANTSATGTIGTISGNTVAVTANAD--LTVG 1235

RESULT 3
US-08-530-198-4
; Sequence 4, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J

```





**THIS PAGE BLANK (USPTO)**

**BEST AVAILABLE COPY**

---

CC in the diagnosis of *C. pneumoniae* infection in mammals. The  
CC nucleic acids and proteins can also be used in the immunization of  
CC mammals, the nucleic acids being particularly useful as DNA  
CC vaccines for effecting *in vivo* expression of antigens. The  
CC vaccines may also prevent atherosclerosis and bronchial asthma,  
CC which are possibly associated with *C. pneumoniae*.

SQ Sequence 918 AA;

Query Match 32.4%; Score 1394.5; DB 20; Length 918;

Best Local Similarity 35.4%; Pred. No. 1.1e-85;

Matches 334; Conservative 139; Mismatches 343; Indels 127; Gaps 18;

```
OY 1 MKIPRLFLI--SLVPTLSMNLGAATTEELASNSFPGTSTSFSSK-TSATDGN 57
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 mssfsflllssslafplmsvsadaadl-tlgsrdsyngdltctetpkaatsdasglt 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 58 YVEKDSVIENVPKTEGQSTSCFRNDAAGDLNFGGFSFTFSNIDATTAAGAAIGSE 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 60 yllgdvslsgqkq-talttscfn--tagnltfignfshfnllstvagvyvant 116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 118 AANKVTLSGFSALFLKSPAST-----VTNGL----- 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 117 aasgltktsfistlmlaaprtltgkaikltgltvfyfsglndqenassengalnkt 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 146 -----GAINVKGNTLS 155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 177 lslgtstrfvalfignssgqgaalyasgdsvlisenagllsfmnsatlsaggaalsegnlv 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 156 ILDNKVLIDNFTSGDGAING--AGS-----LKIANKSLSFIGNSSSTRGAIHRTN 208
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 237 lsnnqnlffdgckatngaidcnkagandpblclsgneslhflntagnsgaalytkk 296
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 209 LTLSSG-GETLEOGNTAPPAAGKGAIAIADSGTSLISDSDIIFEGNTIGATG---TV 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297 lylssgrgvtlsmnkanaapkggialldsgelsiadlgnllfegntltstgspasv 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 265 SHSAIDLGTSKATITLRAAOGHTIFYDPITVWTGTSVADALINSPDPTGDNKEKXTGV 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 357 tnaidlasnakfhlratrignkvlfydpitssgat---dklslnkadagsnlyegylv 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 325 FSGEKLTAEAKDEKRNRTSKLQNVAFKNGTVYLKGDVYLSANGFSODANSKLINDLGT 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 414 fageklseeeelkpnldkctfcgavelaagalvldgvyrvanltlqvegskvmdglt 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 365 LVANTESELTLEINIDSLRNGKKIKLSAATAQDIRIDRPVLAISDESFYONGFLNE 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 474 feasaeyvtlglainldslgdtnkalkataaskdvalspimlvdaqnyehhnlsq 533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 445 DHSYDGIIELDGKDVIYS--ADSRISINAVOSPYGOGKWTINWSTD---DKKATVSWA 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 gqvtp-llelsqgtmcttdlptclpnlnt-nhygyqgqllvwvddataktknatltwt 591
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 499 KOSFNPTAEQEAFLVNLMLGSLFIDVRPQNFIEGTGAPYEKRFVWAGISNLHRSGR 558
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 592 ktygkpnperqplvpnslwgsvfdvrsqslmdrstssslswslwsgiadflhedqx 651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 559 ENOKRFRHVSAGAVVGASTRMGCDTSLSGFQQLPAROKDIFMNTNFATYAGSLRLQHD 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 652 gndrtyrhtsaayaalgggfftasenffnfafqclfygdhllvaknhchvyagamsyrhl 711
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 619 ASLYSVSILLEGGLRELLPYVSKTLPCSFYQOLSYGHTDHRMKTESLPPPTLSTD 678
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 712 gesktlaktlsgn-----sdsipfvfnarfayghdmmktkytyspv----- 755
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 679 HTSMGCVYWAGELGTRVAVENTSGRGFFREYTPPVKVOAVYSRODSFVELAISRDPSDS 738
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 756 ksgwgnatfgiecggaipvasgrswdthpflnlmllyahqndfkengfegrsfgse 815
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 739 HLYNLAIPGLIKLERFAAGQYHVAVAMSPDYCRSNPKCTTLLSNOGSKTKGSNLARQ 798
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 816 dlfnlavpvgikfexfcdsktydlslaypvdvrlndpgcttllmwsdswstcgslsrq 875
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

OY 799 AGIYOAGFSRISGAAAELEFGNPFEMRGSSRSRYNDAGSKIRF 841  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB 876 alivragmhafasnfertsgfvelrjgsrsyaldlggrifg 918  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: October 2, 2001, 03:26:41  
Job time: 30233 sec



```

XX 23-JUN-1997; 97DK-0000744.
PR (BIRK/) BIRKELUND S.
XX (CHRI/) CHRISTIANSEN G.
XX
XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mygind P;
XX WPI: 1999-105610/09.
DR N-PSDB; AAX06823.
XX
XX Species-specific test for identifying mammals infected with
XX Chlamydia pneumoniae - comprises detecting antibodies specific for
XX outer membrane proteins of C. pneumoniae or nucleic acids encoding
XX these proteins
XX
XX Claim 7; Page 63-65; 115pp; English.
XX
XX This polypeptide comprises the novel 97.6 kDa surface exposed
XX protein Omp1 of the human respiratory pathogen Chlamydia
XX pneumoniae. Its amino acid sequence was deduced from DNA (see
XX AAX06823) isolated from a C. pneumoniae expression library. The
XX invention provides 12 novel surface exposed proteins, Omp4-Omp15
XX (see AAM88417-28), and nucleic acid sequences encoding them (see
XX AAX06816-27). A new species specific test is claimed that is used
XX to identify mammals (including humans) infected with Chlamydia
XX pneumoniae. The test comprises detecting antibodies specific for
XX Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
XX membrane proteins, especially by PCR. The proteins are also used
XX in the diagnosis of C. pneumoniae infection in mammals. The
XX nucleic acids and proteins can also be used in the immunization of
XX mammals, the nucleic acids being particularly useful as DNA
XX vaccines for effecting in vivo expression of antigens. The
XX vaccines may also prevent atherosclerosis and bronchial asthma,
XX which are possibly associated with C. pneumoniae.
XX
XX Sequence 930 AA:
SQ

```

```

Query Match 32.7%; Score 1404.5; DB 20; Length 930;
Best Local Similarity 36.9%; Pred. No. 2.4e-86;
Matches 354; Conservative 124; Mismatches 333; Indels 149; Gaps 20;

```

```

OY 1 MKIPRLFLISLVPTLSMSNLGAAT---TEELASNSFDFGTTSTFSFKTSATDGTN 57
DB 1 mltiplkllis--slvtrp1llistatygaadaasptdsddagsgst-fpkstadaangn 57
OY 58 YVFKDSVVIENVPKGTGETOSTSCFKNDAAAGDLNFIYGGFSEFTFSNIDATTAAGAAIGSE 117
DB 58 ylsqgnvyindagk-gtaltygcftc--ctgdlftfykgyfsfnvdagsnagaa-ast 113
OY 118 AANKVTYLSGFSALSLFKPASTVNTGGAIVNKGNLISLDDMKVLIIDNFS----- 169
DB 114 tadkaltfctgnisltiaaqpcltvaagksctissagalnltldngtllfsgvnneannng 173
OY 170 ----- 169
DB 174 alttkltisgntssitfssnaskligalyssaasistngqlvfmnkgetggaly 233
OY 170 -----TGDCGAINC-----AGSLKIANKKSLSFTGNSSSTNG 201
DB 234 faassaltgnslffsgntatdaagkygalycektgeptcltltsgkscitfaensvtg 293
OY 202 GAHTKNTLTLSSGGETLFGGN-TAPTAAGGAIAIADSGTSLISGDSGDIIFEGNTGA 260
DB 294 gatacaghltsaagptlfnmrcgntaagkyataadsgsislsangqcltltgltls 353
OY 261 TG-TVSHSAIDLGTSAKITLALRAAGHTIYFDPITTVGSGTSVADALINSPDTGDNKE 318
DB 354 tsapstlnalylgssakltlnlraagsgslyfydpt-aanttgasvltlnqdpnspld 412
OY 319 YGCTIVFSGEKLTEAKADEKKNRTSKLQNVAFKNGTVVLKGDVYLSANGFSODANSKLI 378

```

```

DB 413 ysgtlvfgskltsadeakaadnftsilkgpllaasgtlalkgveldvngftqegstll 472
OY 379 MDLGTSLVANTSEIETNLNINDSLRNGKKITKLAATOKORIDRPVVALISDESPFQ 438
DB 473 mgpytklkadteaistlkivdlsalegnksvleaganktltlspivfdsgnfy 532
OY 439 NGFLNEDHSYDGL--ELDAGKDIVISADSRSSINAVOSP---YGYOGKWTINM--STDOK 491
DB 533 shltngafqplvftaataadiyl--dalltspvqtrpehygygghweatwadstak 590
OY 492 KATVSWAKOSFNPTAQEAPLVPNNILMGSFIDVRPQNFIELEGAPYEKRPVAGISN 551
DB 591 sgmltwvttygpnperasvvpdsiwasftdlrtltqmtsqanslyqrgjmaasgtan 650
OY 552 VLHRSGERNQKFRHYSGGAVYGASTRMFGCDTLISGFNOLFARDVDYMNNTFAITYAG 611
DB 651 ftkdksgtngatfrhksygyivggsaedfsenlfsvalcqltgykdldlventshyla 710
OY 612 SLRIQHDASLYSVSYLLGEGGLREILPYVS-----KTLPCSFYQOLSYGHTDHHM 663
DB 711 sllyqhratf-----gg-----lmpstfgstldmkdpllnagqlsytykndm 755
OY 664 KTESLPPPPPTLSTDHTSMGCVYVWAGELGTRVAVENTSGRGFFREYTPPVKQAVYSROD 723
DB 756 dcrtytyspa-----qswtlnsgalelgslalylpkeapffgyfptlkfqayvsrq 810
OY 724 SYVELGAIKRDSDSHLYMLATPLGKLEKRRAEQ--YHYVAMYSPPDRCRSPKCTTL 781
DB 811 ntkesgaearatddgdlvcsilpvglrleklsedeknftlsanjdvyrxnprtsrtl 870
OY 782 LSNQGSWTKRKSGLAROGAIVQASGFRSLGAALLEGNGFEGWGRSSRYNDAGSKIRF 841
DB 871 mvsgaswtclcknlraqatlaasagshltlspivelsgeaayelrgsahlnyvndcglrysf 930

```

```

RESULT 14
AAM88418
ID AAM88418 standard; Protein; 928 AA.
XX
AC AAM88418;
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp5.
XX
KW Omp5; outer membrane protein 5; surface exposed protein; antigen;
XX infection; diagnosis; vaccine; atherosclerosis; asthma.
XX
OS Chlamydia pneumoniae.
XX
PN W09858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98WO-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
XX (CHRI/) CHRISTIANSEN G.
XX
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX Mygind P;
XX
DR WPI: 1999-105610/09.
XX N-PSDB; AAX06817.
XX
XX Species-specific test for identifying mammals infected with
XX Chlamydia pneumoniae - comprises detecting antibodies specific for
XX outer membrane proteins of C. pneumoniae or nucleic acids encoding
XX these proteins
XX

```







Db 373 aalnsgnktelraeknhltffdydpiltses--dvikinnsgsagalnpygltlfsq 430  
QY 328 EKLTEAEAKDEKNTSKLLQNVAFKNGTAVLKGDVVLASANGFSDASXKIMDGTSLVA 387  
Db 431 etltaedlkvaadkksftcpvpslsgkllllqkyvltessfsqgaagllgmstlslst 490  
QY 388 NTFSEIETNEINIDSLRNKKIKLSAATAOKDIRDPVVALISDSFYONGFLNEDHS 447  
Db 491 tagstftnlglnvdslylqpslatakasnkylvsgknlldiegnlyeshmfsthdq 550  
QY 448 YDCLLELDACKDIVISADSSINAV-----QSPGYOGKMTINKSTD---DKKATVSWA 498  
Db 551 f-sllkltvdadvtdnvdisslllpvpaedpaseygfqgavvntdcatntkeatwtc 609  
QY 499 KQSFNPTEAEQAPLVPLMLMGSPFIDVRPQNFTELGTEDGAPYEKRFVAVASINVLHRSGR 558  
Db 610 ktgfvpspekalsavcvtlwgvtfdlrsllqylveigtatgmehqyglvssmtlflhtqd 669  
QY 559 ENORKEFHVSGAVGAVGASTWMPGCDTLGLGPAOLFARDKDYFMNTNPAKTYAGSLRQHD 618  
Db 670 enrkgtfhtsgylyvgsahtpkddltfafaclhfardkcclflahmstrtyggtlffkhs 729  
QY 619 ASLYSVSILGEGGKREILLPVYSKTLPCSPFYGQLSYGHTDHRMKTG--SLPPPTLS 676  
Db 730 htlqpgnyllrlgfrakfsesalekfpceipaldvqvasfshsdmethylslp----- 782  
QY 677 TDHTSMGYVMAGELGRVAVENTSGRGFFREYTPFKVQAVYSRODSFVELASRDFS 736  
Db 783 esegswsneclaggldlfpvlshpnlkltltpqmkvemyvsgnftessdgtgfs 842  
QY 737 DSHLYNLAIPLGIT-LEKRAEQY-YHVVAWSPDVCRSNPCKTTLTLLSQSGWKTGKN 794  
Db 843 lgrllnslpvgakftvgdlygdslytalsgffsvdyrmpgstaltvmpdewkrltgrn 902  
QY 795 LAQAQGVQASGFRSLGAAAELEFGNCFEMRGSSRSYNDAGSKITR 841  
Db 903 lsrqatflrsgsnnyvyncelifghyamelrsgsrnyvndvgtklrf 949  
RESULT 10  
AA35052  
ID AA35052 standard; Protein; 930 AA.  
AC AA35052;  
XX  
XX AAY35052;  
DT 13-SEP-1999 (first entry)  
XX  
XX Chlamydia pneumoniae surface exposed polypeptide.  
DE  
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.  
XX  
XX Chlamydia pneumoniae.  
OS  
XX  
XX MO9927105-A2.  
PN  
XX  
XX 03-JUN-1999.  
PD  
XX  
XX 20-NOV-1998; 98WO-1B01890.  
PF  
XX  
XX 04-NOV-1998; 98US-0107078.  
PR  
XX 21-NOV-1997; 97FR-0014673.  
XX  
XX (GEST ) GENSET.  
PA  
XX Griffais R;  
PI  
XX  
XX MPI; 1999-357842/30.  
DR  
XX  
XX Genome sequence of Chlamydia pneumoniae  
PT  
XX  
XX Page 940-942; Disclosure; 1912pp; English.  
PS

XX  
CC AAY34584-y35879 represent the proteins encoded by all the open reading  
CC frames in the complete genome (see AAY91990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAY34584-y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.  
CC  
CC  
XX  
SQ Sequence 930 AA:  
Query Match 32.9%; Score 1413.5; DB 20; Length 930;  
Best Local Similarity 37.0%; Pred. No. 6e-87;  
Matches 355; Conservative 144; Mismatches 332; Indels 149; Gaps 20;  
QY 1 MKPIPLRLISLVPYTLMSNLGAAT---TELSASNSPFGCTSTFSKTSADGTN 57  
Db 1 mkplplhllis--slvtpllslatygadaslsptdsfdaagst-fpkstadanlgn 57  
QY 58 YVFKDSVVIENVPKGTGCTGCTSCFKNDAAADLNFLLGGGSEFTFSNIDATTAAGAAIGSE 117  
Db 58 yvlsgvnyindagk-gtalgcctfe--ttgdltfkgysfsvntvdsgnagaa-ast 113  
QY 118 AANKVTYLSGFSALSPFKSPASTVTNGLGAINVKNLSLDNDKVLIDNFS----- 169  
Db 114 tadkaltftgfnslfstaapgtlvaagkctslsagalnldngtllfsgvnsneannng 173  
QY 170 ----- 169  
Db 174 altaktlsigntssltftsnsakllggaalysaaasigntqvlvfmmnketggagly 223  
QY 170 -----TGDCGAINC-----AGSLKATANKSLSPFGNSSSTRG 201  
Db 234 feaassltgnsslffsgntatdaagkgyalcetgetpbltllsgnkslfaenssvtgg 233  
QY 202 GAHTKKNLTVSSGGETLFOGN-TAPTPAAGKGAIAINDSGTSLISGDSGDIIFEGNTIGA 260  
Db 294 galcabgldisaagpctlfsmrcngntaagkgaladagslalsanqgdlftlgnlts 353  
QY 261 TG--TVSHSAIDGTSKKTALRAAGCHTYFYDPITYGSTVAADALNNSPDGDNKE 318  
Db 354 tsapstclnaillgssakltnlraaqgslytdpl-asnltgasadvltlnqpsnpld 412  
QY 319 YTGTVFSGEKLTEAEAKDEKNTSKLLQNVAFKNGTAVLKGDVVLASANGFSDASXKLI 378  
Db 413 ysgtlvrsgekladeakadnftslkqplalaagtlakgnveldvngftlqegstll 472  
QY 379 MDLGTSLVANTESIELTNEINIDSLRNKKIKLSAATAOKDIRDPVVALISDSFYONG 438  
Db 473 mmpgkrlkadteaistklvlvdlsalegnksvalaagankltltsplvfqdsagngfye 532  
QY 439 NGFLNEDHSYDGL--ELDAGKDIVISADSSINAVOSP---GYOGKMTINK--STDDK 491  
Db 533 shclnqafclpvlfaataasadyi--dallspypcphephygyghweatvadstak 550  
QY 492 KATVSNAKQSFNPTEAEQAPLVPLMLMGSPFIDVRPQNFTELGTEDGAPYEKRFVAVASIN 551  
Db 591 sgtmctwtltygnpnperrassvypdsiwaastfdtrtqglmtsganslyqgrglvaagtan 650  
QY 552 VLHRSGRNORKEFRHVSAGAVGASTWMPGCDTLGLGPAOLFARDKDYFMNTNPAKTYAG 611  
Db 651 ffhkdksgtnqaftrhksylyvgsaeedfeenlfsvalcqlfgkdxdlflventshnlya 710  
QY 612 SLRLQHDASLYSVSILGEGGKREILLPVYS-----KTLPCSFYQGLSYGHTDHRM 663  
Db 711 slylqhnaf-----lg-----lmpsfsgsltdmlkdplllnaqslvysykndm 755  
QY 664 KTESLPPPTLSTDHTSMGYVMAGELGRVAVENTSGRGFFREYTPFKVQAVYSROD 723

```

QY 1 MKIPLELLISLVPTLSMNLGAATTEELASANSNDFGTTSTTSFSSKTSATDGTNYVF 60
DB 1 mktslpwlvssvlafs-chlqslaneellspddsfngidsqftfprksa-----tlysl 55
QY 61 KRSVYIENVPKGTGEQSTSCFRNDAAGDLNFLGGGFSTFESNIDATWTASGAIGSEAN 120
DB 56 tgdvffye-pgkgtprlsdcfkg--ttdnltflngshstlfgfidagthagaa-astan 111
QY 121 KNTVLSGFSALSFELKSPASTVNGICAINVKGNLSLDDKVLIDNFTSGGGAINCAG 180
DB 112 knltsgfslsfdspsctvtlqgltssagvnlentkrklvagnfstadggaikgas 171
QY 181 ----- 180
DB 172 fltltsgdalfsmnssstkgaiaatagariantgyvrfisnlastsgaiddgstsl 231
QY 181 -----SLKIANKSLSPFGNSSSTRGGAHTKNTLT 210
DB 232 lsnkflyfegnaakttggaicntkassgspellissnktliffasvnaetsggaiahakkia 291
QY 211 LSSGGETLFOGWTAPTAAGKGAIAIADSGTSLISGDSGDIIFEGNTI--GATGVSHS 267
DB 292 lssggtfeflrnvsasaprkgaaisdaagselsaetgnltfvnltlttsgtdpkrn 351
QY 268 AIDLGTSAKITPLRAAOGHTIIFYDPITVGTSTVADALINISPTDGNKEXTGTIVFSG 327
DB 352 alnigsnogfteliraaknhtlffypdtsegtss--dvklingsagaalmpqgltlffsg 409
QY 328 EKLTEAEAKDENRRIKSLONAVAFKNGVYLKGDVYLSANGSODANSKLMDLTSLVA 387
DB 410 eeltdeklvadhklsfqtgpylsqgklllqkyvlesfestsagaeslllmgndstlct 469
QY 388 NRESIELFNLLEINISLRNGKIKLSAATAQKDIRDRPVVAISDESRYOGNFIEDHS 447
DB 470 tsgstltnlgnlvsalslqtkpvsitakgasnkvlvsqklnlidlagnlyeshmfshdq 529
QY 448 YGGLLELDAGKDIVISADSRISINAV---QSPYGYQKWTINWSTD---DKKATVSA 498
DB 530 f-sllktlvdadvtdnvdnssilpvaedpnsegyfqqgwmvnmwltatnckeatatvt 588
QY 499 KOSFNPTAQEAPLYPNLLMWSFIDVRRQNTLEGTBEGAPKEKRFVWAGISNVLHRSR 558
DB 589 kctgfpvperksalvcnltwgvftdlrslqglveigtatmekkgfivssmtnflhtkgd 648
QY 559 ENOKRFRHSGCAVVGASARMFGCDPLSLGFAQLFARDQDYMNTMFAKTYAGSLRLHD 618
DB 649 enrkgftrhsqyvlvgssenhcpkddlftafchlfarddcfiahnstlygqclffkhs 708
QY 619 ASLYSVASILLGEGGLRETLRPVSKTLPCSPYQQLSYGHTDHMKTE--SLPPPPPLS 676
DB 709 hclqpnqylrlgrakfseesalekfprelplaidvqvsfshsnmethyslp----- 761
QY 677 THTTSMGCVWAGELGTRAVENTSGRPFREYTPPVKQAVYSRODSVEIGALSRDS 736
DB 762 esegsswneclagylqldlpfvlsnphlffklfpkmkewmyvqnsfiesssdgrfhs 821
QY 737 DSHLNLAIPLGLIK-LEKRFAYOY-YHVVAWYSPDVCRSNPCTTILISQOSMWTKGSN 794
DB 822 lgrllnlsipvaktvgdldgsyctydlsgsfvsvdyrnnpstatlvmnspsaklrign 881
QY 795 LARQAGIVQASGFRSLGAALFNGFGEWGRSSRSRYNDAKSIKIF 841
DB 882 lartqflllrgsnnyynsnceelfghyamelrgsrnyndvgtklrft 928

```

```

XX Chlamydia pneumoniae cellular envelope protein.
DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
OS Chlamydia pneumoniae.
XX
PN W09927105-A2.
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-1B01890.
PR 04-NOV-1998; 98DS-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST ) GENSET.
PI Griffais R.
XX WPI: 1999-357842/30.
XX
Genome sequence of Chlamydia pneumoniae
Page 947-949; Disclosure: 1912pp; English.
XX
AA134584-135879 represent the proteins encoded by all the open reading
frames in the complete genome (see AA134584-135879) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
frames of the C. pneumoniae genome (see AA134584-135879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
SQ Sequence 949 AA:
Query Match 33.3%; Score 1429.5; DB 20; Length 949;
Best Local Similarity 35.8%; Pred. No. 5,2e-88;
Matches 339; Conservative 147; Mismatches 336; Indels 125; Gaps 15;
QY 1 MKIPLELLISLVPTLSMNLGAATTEELASANSNDFGTTSTTSFSSKTSATDGTNYVF 60
DB 1 mktslpwlvssvlafs-chlqslaneellspddsfngidsqftfprksa-----tlysl 76
QY 61 KRSVYIENVPKGTGEQSTSCFRNDAAGDLNFLGGGFSTFESNIDATWTASGAIGSEAN 120
DB 77 tgdvffye-pgkgtprlsdcfkg--ttdnltflngshstlfgfidagthagaa-astan 132
QY 121 KNTVLSGFSALSFELKSPASTVNGICAINVKGNLSLDDKVLIDNFTSGGGAINCAG 180
DB 133 knltsgfslsfdspsctvtlqgltssagvnlentkrklvagnfstadggaikgas 192
QY 181 ----- 180
DB 193 fltltsgdalfsmnssstkgaiaatagariantgyvrfisnlastsgaiddgstsl 252
QY 181 -----SLKIANKSLSPFGNSSSTRGGAHTKNTLT 210
DB 253 lsnkflyfegnaakttggaicntkassgspellissnktliffasvnaetsggaiahakkia 312
QY 211 LSSGGETLFOGWTAPTAAGKGAIAIADSGTSLISGDSGDIIFEGNTI--GATGVSHS 267
DB 313 lssggtfeflrnvsasaprkgaaisdaagselsaetgnltfvnltlttsgtdpkrn 372
QY 268 AIDLGTSAKITPLRAAOGHTIIFYDPITVGTSTVADALINISPTDGNKEXTGTIVFSG 327

```

CC membrane proteins, especially by PCR. The proteins are also used  
CC in the diagnosis of *C. pneumoniae* infection in mammals. The  
CC nucleic acids and proteins can also be used in the immunization of  
CC mammals, the nucleic acids being particularly useful as DNA  
CC vaccines for effecting *in vivo* expression of antigens. The  
CC vaccines may also prevent atherosclerosis and bronchial asthma,  
CC which are possibly associated with *C. pneumoniae*.

50 Sequence 928 AA;

Query Match	33.3%	Score 1429.5;	DB 20;	Length 928;
Best Local Similarity	35.8%;	Pred. No. 5e-88;		
Matches 339;	Conservative 147;	Mismatches 336;	Indels 125;	Gaps 15;

QY	1	MKPIRELLISLIVPTLMSNLGAATTHELSANSRPGCTTSTNSESSEKSSADGNTYVF	60
Db	1	mktslpwlvssvtaifs- ch,qslsaneellspdds frngaidsglfrpktsa- ---tlvs1	55
QY	61	KDSVVIENVPRTGTOSTSCFPKNMAAGDNLFLGSGFSFTFSNIDATATAGAAIGSEAN	120
Db	56	tdgvfrfyepkkygplsdscfkq- ---tdnltflgnngslstfgfidagthaae- astan	111
QY	121	KTVTLISGFSALSFLKSPASTVTVNGLAIMVKNLSLLDNDKVLIDQNFSTGDSGALNCAG	180
Db	112	knltfsgslstfadspsltvtggtlssagynlenirkllvagnfstadgalkgas	171
QY	181	-----	180
Db	172	flftgsgdalflsnssslkgyaataatagarianncyvflfnlaistsgaiddegti	231
QY	181	-----SLKTAANKSLSEFIONSSSTRGAIHRTNLT	210
Db	232	lsmnkflyfegnnaaktcgtgaicnckasgspehlismnkllifanvetsgaihakkla	291
QY	211	LSSGGTLEPGCNTAPTPAAGKGAIALADSGTSLISGSDIIEBGNTI- ---GATGVSHS	267
Db	292	lssggtfelflnnvssatpkgaistsdasglsistsaetgnltfyrnlitlttgstdcpkn	351
QY	268	AIDIGTSKIRALAAAGCHTTFYFDPTTVYNGSTVSAAALMINSDDTQDNKNEYGTIFYSG	327
Db	352	ainfgsnqkftelnaaknhtlfrpdlitsegts- dvlknngsagalnpypgqrlfllsg	409
QY	328	EKLTEAEAKDKENRTSKLQVNAKNGTIVVLKGVVLSANGFSQDANASKIMDGLSTSLVA	387
Db	410	etlradelkvadnllksftcgtvpslsgklllqkvltlescsfgeagsllgmssqtlstc	469
QY	388	NTESELTNLEINIDSLRNGKKIKLSAATQKDIRIDRPVVLAISDSFQNGFLNEDHS	447
Db	470	tagaiftlnlgnidvdsjglkqpsvltakgasnkvlyvsklnlidlagnlyeshmfshdql	529
QY	448	YDGLIELDAGKDIYISADSRKINV- ---QSPVGYQSKMTINKSTP- ---DKKATVSM	498
Db	530	f-sllklivdadvndlvnslslfpyrpaeðpnseyigqgvnmvnttcatnctaeatwt	588
QY	499	KQSENPTEAEAPLVPVNLMSFIDVRRPQNFILGTGEGAPYEKREKRVAGISVNLHRSGR	558
Db	589	ktgfrvperpsalsavcntlmgvfcdirsldqiveigdtgmehqvgfivssamtnflhtgd	648
QY	559	ENQKRFHVSGAAYVGASTRMPGGDTLSLCPAQLFANDKDYFMNTEFAKTVAGSLRLQHD	618
Db	649	enrkyfntsggyyiggsahnpkddltftafcnlfardckdcflahnnstrcygglffkhs	708
QY	619	ASLXSVSIIILGEGELNELILPYVSKTLPCSFYQQLSYGTHDRMKRE- ---SLPPPTTSL	676
Db	709	htlpqnyliragrakfesaalekfrpeipalaldqvafsfshdsndmehyslp- ---	761
QY	677	TDHSTMGCVWAGELGTRVAVENTSGRGFRFVPEFKVOAYVSRQDSFELASRDFS	736
Db	762	esegsvsneclaaagldpfrvlsnbnpllktflfgpmkvemvyysqnsflessdsdgrgfs	821
QY	737	DSHLVNLAIPLIGIK-LEKRFRAEØ- YHVAVAMYSDVCRSNPCKCTTLLISNOGSWKTKGSN	794

Species-specific test for identifying mammals infected with Chlamydia pneumoniae - comprises detecting antibodies specific for outer membrane proteins of C. pneumoniae or nucleic acids encoding these proteins

Claim 7, Page 53-55; 115pp; English.

This polypeptide comprises the novel 90.0 kDa surface exposed protein omp8 of the human respiratory pathogen Chlamydia pneumoniae. Its amino acid sequence was deduced from DNA (see AAX06820) isolated from a C. pneumoniae expression library. The invention provides 12 novel surface exposed proteins, Omp4-Omp15 (see AAM88417-28), and nucleic acid sequences encoding them (see AAX06816-27). A new species specific test is claimed that is used to identify mammals (including humans) infected with Chlamydia pneumoniae. The test comprises detecting antibodies specific for Omp4-Omp15 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used in the diagnosis of C. pneumoniae infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with C. pneumoniae.

Sequence 928 AA:

Query Match 36.1%; Score 1553; DB 20; Length 928;  
Best Local Similarity 39.0%; Pred. No. 2.4e-96;  
Matches 367; Conservative 121; Mismatches 314; Indels 140; Gaps 20;

```

OY 15 TMSNMLGAATTEELSNFDTSTTSFSSKTSATDGNVYFKDSVLENTPKGE 74
DB 12 tfaipmliaetvldasafdg-nkngnsvresgedagtytkfgnvtlenipgqt 70
OY 75 TOSTSCFNDAAGDLNFGGFSFTFSNIDATATASGAIGSEANKRYTLLSGFALSFL 134
DB 71 attkacfm--tkgltftgngsnllftqvadgtvagaavnsvvdkettffgfslsfti 128
OY 135 KSPASTVINGLCAINVK-GNLSLDNDKVLIDNFSSTDGGAINC----- 178
DB 129 aspgsiltgkavscstgslkfdkvnslfkskfstdgaltaktlstgtmsaifls 188
OY 179 -----AGSLKIANNKSLSFI--- 193
DB 189 entsskggaigtcdaltltngngevstsdtsdsgaaitfeasvltisnakaifidnk 248
OY 194 -----GNSSSTRGGAIRHKNLTLSSG 214
DB 249 vtgaassrtgdmssgaicaaykstctkvtltngnmlfsmntstagaaiyykklelnasg 308
OY 215 GFTLRF-----QGNAPTPAGKGAIAIADSGTSLISGDSGDIIFEGNTIGAN--GTVSHS 267
DB 309 gltlfsmvnggtap----kysaiaiedsgelsisadsqdivlfgntvstptp--rrs 363
OY 268 AIDLTGSAKITALRAOGHTIYFVDPIYVWSTSVADALNINSPDTGKNETGTIVISG 327
DB 364 sldlgtscakmaltisaagraalyqdpilcgsstclvdtkvnepradaalqlygnliltg 423
OY 328 EKLTEAEAKDEKNTSKLQNVAFKNGTVLVKGDVVLISANGFSQDANSKLINDLCTSL-V 386
DB 424 eklseteadaadsknltskllpyvtlsgtflskhyvltlgtatgdaerlendvgttletp 483
OY 387 ANTESIELTNEINIDSLRNKGKIKLSAATYAKOKDIRDPVYLAISDSFYONGFLNEDH 446
DB 484 adtsli--nnvinissidgakkakleakatsknltlsgtllldpigtlyfhenhlmprp 541
OY 447 SYDGLIELDAGKDIYISDSRS--INAVOSPYGOGKW--TIMMSTD--DKKATYSMAKQSFN 503
DB 542 syd-lllekaasgtvscavtpdpjngkfhygyggtwpiyvgtaastatcltnwkttyi 600
OY 504 PTAEOEAPLVPNILMGSEFIDVRFQNFIELGTGAPAEKREFWAGISVNLHRSGRENOR 563

```

```

DB 601 pperlsgslvpslnmafidslshlymetaneglgqdratfcaglsnfihkdstktrrg 660
OY 564 FRHVSAGAVGASTRMPGGDTLSLGFAGQLFARDKIDFMNNTAKTYAGSLRQHDASLXS 623
DB 661 frhlsygyvlggnlhtcsqklisaaficqlfgrdrdyfvakngltyvgflllyghmety 720
OY 624 VYSILSGEGAREILLPYVSKRLPCSFQGLSYGHTDHMKTESLPPPPPTSTHTSMG 683
DB 721 lpc-----klrpsalsvptelpvlfsngnlytchdnolkky-----tlyptvksvsg 769
OY 684 GYVWAGELGTR--VAVENTSGRGFFREYRPEYKVOAVYSRQDSFVELGASIRFSDSHLY 741
DB 770 ndsfalelgydrpicldesa---lfeqymfmlklfyvahqgfieqgtrearefssrlv 826
OY 742 NIATPLGKLEKR--FAEQYHVAMYSFQVCRSNPKCTTLLISQGSWKTKGSLARQA 799
DB 827 nlalplgtrfdkesdcgdtlyltlgytvdvtrsnpdccttlrtisgdswwktfgtnlarqa 886
OY 800 GIVQASGFRSLGAAPLEKGNFGEWRGSSRSYVNDAGSKIRF 841
DB 887 lviraghnfcfnfcafsqsfelrsgsrnyvndlgakqyf 928

```

RESULT 7  
ID AAM88417 standard; Protein; 928 AA.

AC AAM88417;

DT 26-APR-1999 (first entry)

DE Chlamydia pneumoniae surface exposed protein Omp4.

KW Omp4; outer membrane protein 4; surface exposed protein; antigen; Infection; diagnosis; vaccine; atherosclerosis; asthma.

KM Chlamydia pneumoniae.

PN W09858953-A2.

PD 30-DEC-1998.

PF 19-JUN-1998; 98WO-DK00266.

PR 23-JUN-1997; 97DK-0000744.

PA (BIRK/) BIRKELUND S.

PA (CHR1/) CHRISTIANSEN G.

PI Birkelund S, Christiansen G, Knudsen K, Madsen A;

PI Mygind P;

DR WPI; 1999-105610/09.

DR N-PSDB; AAX06816.

PT Species-specific test for identifying mammals infected with

PT Chlamydia pneumoniae - comprises detecting antibodies specific for

PT outer membrane proteins of C. pneumoniae or nucleic acids encoding

PT these proteins

PS Claim 7, Page 40-42; 115pp; English.

CC This polypeptide comprises the novel 98.9 kDa surface exposed

CC protein Omp4 of the human respiratory pathogen Chlamydia

CC pneumoniae. Its amino acid sequence was deduced from DNA (see

CC AAX06816) isolated from a C. pneumoniae expression library. The

CC invention provides 12 novel surface exposed proteins, Omp4-Omp15

CC (see AAM88417-28), and nucleic acid sequences encoding them (see

CC AAX06816-27). A new species specific test is claimed that is used

CC to identify mammals (including humans) infected with Chlamydia

CC pneumoniae. The test comprises detecting antibodies specific for

CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer

PN	WO200026237-A2.
XX	
PD	11-MAY-2000.
XX	
PE	29-OCT-1999; 99WO-GB03579.
XX	
PR	29-OCT-1998; 98US-0106070.
PR	01-MAR-1999; 99US-0122066.
PR	27-OCT-1999; 99US-0428122.
XX	
PA	(CONN-) CONNAUGHT LAB LTD.
XX	
PI	Murdin AD, Oomen RP, Dunn FL;
XX	
DR	WPI: 2000-365569/31.
DR	N-PSDB: AAA27021.
XX	
PT	Novel Chlamydia 98 kDa putative outer membrane protein antigen, used
PT	for vaccination and protection against Chlamydia infection -
XX	
PS	Claim 6; Fig 1; 93pp: English.

CC The present sequence is the 98kDa putative outer membrane protein from  
CC Chlamydia pneumoniae. The genomic sequence was amplified using two  
CC PCR primers. The 5' primer contains a NotI restriction site, a ribosome  
CC binding site, an initiation codon and a sequence close to the 5' end of  
CC the 98kDa putative outer membrane protein coding sequence. The 3' primer  
CC contains the sequence encoding the C-terminal sequence of the putative  
CC outer membrane protein and a BstXI restriction site. The stop codon was  
CC excluded and an additional nucleotide was inserted to obtain an in-frame  
CC C-terminal fusion with the Histidine tag. The PCR product was cloned  
CC into a eukaryotic expression vector (pC4-Myc-His) by restricting both  
CC the vector and the PCR product with NotI and BamHI and performing a  
CC ligation reaction. This expression vector was injected intramuscularly  
CC and intranasally into mice, which were subsequently inoculated with  
CC Chlamydia pneumoniae. The chlamydial lung titers of the immunised mice  
CC were lower than those of the controls. Thus the 98kDa putative outer  
CC membrane protein can be used as a vaccine to provide protection against  
CC Chlamydia infections, especially Chlamydia pneumoniae infections.  
CC The present polypeptide may also be administered orally to treat  
CC Chlamydia infection.

**SQ Sequence 928 AA;**

Query Match	36.4%	Score 1564	DB 21	Length 928
Best Local Similarity	39.2%	Pred. No. 4.3e-97		
Matches 369	Conservative 121	Mismatches 312	Indels 140	Gaps 20

```

OY      15  TLMSNLGAATTEELASNFSDTSTSTSSKTSSTADCTNVFDSVYIENVPATGE  74
        | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db      12  lfalfpmsitelvclsdasafdg-nkngmslvesesgedatylfifgnvllnlpigt  70

OY      75  TQSTSCFKNDAAGDLNLFGLGGFFSTFSSNIDATFASGALISEANKVTWLSGFSASFL  134
        | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db      71  altsccfn--tkgdltltngnmlfqtvdagclvgaavrnsvvdcstltigfssifl  122

OY      135  KSPASTVTNGALINVK-GNLSLLDNKVLIIQDNFSTGCGAINC-----  178
        | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db      129  aspgsiltgavscstgsislkvnslfkskfstdngalatlaksltgtmualfs  188

OY      179  -----AGSLKIANKSLSFT--  199

Db      189  entsskkgaiqtdaltltngvgevsfdntssdsgaalftaesvatslsmaksvifdnk  248
        | : : | : : | : : | : : | : : | : : | : : | : : | : : |

OY      194  -----GNSSSTRGCAIHTKMLTSSG  214
        | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db      249  vtgasssttgmsggaicaayktsctckvcltltngnmlfsmstctettagaayvkkleiasg  308

OY      215  GETLF-----QGNTAPRAAGGGAIAIADSTGLISDSDGIIIEGWTIGAT--GYSHS  267
        | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db      309  gltlftrsvngvgcap---kgsaialedsgelslssadgvlvrlgtvtsttsgt--nrs  365

```

[illegible]

RESULT	6
AAW88421	
ID	AAW88421 standard; Protein; 928 AA

AC AAW88421;

DT 26-APR-1999 (first entry)

DE Chlamydia pneumoniae surface exposed protein Omp8.

KW  
OMP8; outer membrane protein 8; surface exposed protein; antigen;  
infection; diagnosis; vaccine; atherosclerosis; asthma.

OS Chlamydia pneumoniae.

PN W09858953-A2.

PD 30-DEC-1998.

PF 19-JUN-1998; 98WO-DK00266.

PR 23-JUN-1997; 97DK-0000744.

PA (BIRK/) BIRKELUND S.

XX

PI Mygind P;

DR WPI; 1999-105610/09.

XX

AAW88421

XX

XX

22

XX

KW infection; diagnosis; vacc

OS Chlamydia pneumoniae.

PN W09858953-A2.

PD 30-DEC-1998.

PF 19-JUN-1998; 98WO-DK0026

PR 23-JUN-1997; 97DK-000074

PA (BIRK/) BIRKELUND S.

XX

PI Mygind P;

DR WPI; 1999-105610/09.

XX

```
OY 209 LTLSSGGETLEFGONTAPPAAGKGAIAIADSGTSLISGSDIIFE--GNTIGANTVSH 266
DB 113 -----gk-gaikltg-----lvfesign----- 130
OY 267 SAIDGTSKAKTALAAAGCHTTFYFDPTVTWGSISVADALNINSPDTGDNKRYTGTYIFS 326
DB 131 -----ldplvtgstsvadalininspdtgdnkeytgtlvyfs 166
OY 327 GKKLLEAEKDEKNTSKLQNVAFKNGTVYLKGDVVLISANGESODANSKILMDGTSLV 386
DB 167 gkllleaeakdekntskllqnvafkngtlvylkgdvvlsangfsgdanskilmdlgtslv 226
OY 387 ANTESIELTNLEINIDSLRNGKIKLSATAQKDIRIDRPVVLAIISDSFYONGFLNEDH 446
DB 227 antesieltnlelnidslrngkiklsataqkdiridrpvvlaisdesfyngflnedh 286
OY 447 STDGILELDAGKDIYISADSRISINAVOSPYGQKWTINMSTDOKKATVSWAKOSFNPTA 506
DB 287 sydgilleldagkdviysadsrsidavspygqkwtnmstdokkatvswakgsfnpta 346
OY 507 EQEAPLVNRLMGSRFDVRRPQNFTELGEGAPYKRRFVAGISVHLRSGRENQRKRRH 566
DB 347 eqeaplvnrlmgstfdvrrpqnftelgtegapyrkrrfvagisvhlrsgrenqrkrrh 406
OY 567 VSGGAVVGASTRMPGCDLSLGFALFARDKDYFMNTNFAKTYAGSLRLQHDASTSYVS 626
DB 407 vsggavvgastmpgdcslsgfalardkdyfmntnfaakyagslrlqhdsstsyvs 466
OY 627 ILGEGGLREILLPVYSTLPSCSYGQLSYGHTDHMKTESLPPPPPTLSTDHTSMGCVY 686
DB 467 ilgegglrellpvystlpcsfygqlsygthdhrmkteslppppptlsthtsmgvy 526
OY 687 WAGELGTRAVENTSGRGFRFRTPEFVVOAVYSRODSFVEIGALSRODSHLYNLAIAP 746
DB 527 wagelgtraventsgrgfrfretpefvvoavysrodsfveigalsrdsdshlynlaiap 586
OY 747 LGIKLEKREAEQYHVAVMSPDV 770
DB 587 lgiklekreaeqyhhvavmypadv 610

RESULT 4
AAY34609
ID AAY34609 standard; Protein: 483 AA.
XX AAY34609;
AC
XX 13-SEP-1999 (first entry)
DT
XX Chlamydia pneumoniae transmembrane protein sequence.
DE
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KM sinusitis; purulent otitis media; erythema nodosum; paronychia;
KM vaccine; neutralising epitope.
XX
XX Chlamydia pneumoniae.
OS
XX MO9927105-A2.
PN
XX 03-JUN-1999.
PD
XX 20-NOV-1998; 98WO-1B01890.
PF
XX 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97ER-0014673.
XX
XX (BEST ) GENSET.
PA
XX Griffiths R;
PI
XX WPI. 1999-357842/30.
XX
XX Genome sequence of Chlamydia pneumoniae
PT
```

```
XX Page 630-631; Disclosure: 1912pp; English.
PS
XX
CC AAY34584-135879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAY91900) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or paronychia. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-135879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
CC
SQ Sequence 483 AA;

Query Match 58.1%; Score 2499; DB 20; Length 483;
Best Local Similarity 99.2%; Pred. No. 1,6e-160;
Matches 479; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 359 KGDVVLISANGESODANSKILMDGTSLVANTESIELTNEINIDSLRNGKIKLSAATAQ 418
DB 1 kgdvvlsangfsgdanskilmdlgtslvantesieltnlelnidslrngkiklsataq 60
OY 419 KDIRIDRPVVLAIISDSFYONGFLNEDHSYDGLLELDAGKQIVISADSRISINAVOSPYG 478
DB 61 kdiridrpvvlaisdesfyngflnedhsydgilleldagkdviysadsrsidavspyg 120
OY 479 QCKWTINMSTDOKKATVSWAKOSFNPTAEOEAPLVNRLMGSRFDVRRPQNFTELGTEGA 538
DB 121 qckwtlnmstdokkatvswakgsfnptaeeaplvnrlmgstfdvrrpqnftelgtega 180
OY 539 PYEKRFVWAGISNVHLRSGRENQRKRRHVSGAVVGASTRMPGCDLSLGFALFARDK 598
DB 181 pyekrfvagienvhlrsgrenqrkrrhvsaggavvgastmpgdcslsgfalardk 240
OY 599 YFMNTNFAKTYAGSLRLQHDASTSYVSTLLGEGGLREILLPVYSTLPSCSYGQLSYGH 658
DB 241 yfmntnfaakyagslrlqhdsstsyvslllegglrellpvykclpccsfygqlsygh 300
OY 659 TDHRMKTESLPPPPPTLSTDHTSMGCVWAGELGTRAVENTSGRGFRFRTPEFVVOAV 718
DB 301 tdhrmkteslppppptlsthtsmgvywagelgtraventsgrgfrfretpefvvoav 360
OY 719 YSRQDSFVELGAISRQDSHLYNLAIPLGILKLEKREAEQYHVAVMSPDVCRSNPKCT 778
DB 361 ysrqdsfvelgaisrdsdshlynlaipllgiklekreaeqyhhvavmypadvcrsnpkct 420
OY 779 TLLSNQSGWKTGKSNLROAGITVAGSGFRSLGAAAEELFGNFGFMRGSSRSYNYDAGSK 838
DB 421 tllsnqsgwktkgsnlraqagivqagisrgaaaelfgnfgfmrsgssrsynvdagsk 480
OY 839 IKF 841
DB 481 ikf 483

RESULT 5
AAY94327
ID AAY94327 standard; Protein: 928 AA.
XX AAY94327;
AC
XX 11-AUG-2000 (first entry)
DT
XX Chlamydia pneumoniae 98KD putative outer membrane protein.
DE
XX Chlamydia pneumoniae 98KD putative outer membrane protein.
KM Chlamydia; antigen; vaccine; infection; outer membrane protein.
XX
XX Chlamydia pneumoniae.
OS
XX
```



CC the presence of Chlamydia nucleic acids in samples (and therefore  
CC diagnose infections) and the proteins may be used as antigens for the  
CC production of antibodies that may be used to detect Chlamydia proteins  
CC in samples (e.g. via enzyme linked immunosorbant assay (ELISA)).

XX  
SQ Sequence 841 AA:

Query Match 99.4%; Score 4271; DB 21; Length 841;  
Best Local Similarity 99.4%; Pred. No. 1.3e-279;  
Matches 836; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 MKIPLRLLSLVPTLSMSNLGAATTEELASNSPFGTTSTTSFSSKTSATDGTNVF 60
DB 1 mkplrlslslvptlsmsnlgaatteeelsaansfgtststssktsatdgtnyvf 60
QY 61 KDSVYIENVKRGTEGOSTSCFRKNDAAAGDINFLGGFSPTFSNIDATTAGAAIGSPAAN 120
DB 61 kdsyvienvkrgtegstscfrkndaaagdinflyggfsptfsmidattagaaigspaan 120
QY 121 KTYTLGSEFSLSPFKSPASTVNTNGAIINVKGNLSLDNDKVLIDNFPSTGDCGAINCAG 180
DB 121 ktytlgsefslspfkspastvntngaiinvkgnlsldndkvliqdnfstgdcgaaincag 180
QY 181 SLRIANKKSLFIGNSSSTTGAIHTKNTLTSGGETLFOGNTAPTAAGGAIAIDSG 240
DB 181 slriankkslfignsssttgaihtkntltsggetylfgnaptagaigaiadsq 240
QY 241 TLTISGDSGIIIEBNTIGATGVSHSAIDLGTSAKITLRAAAGHTIYEDPIYTGST 300
DB 241 tltisgdsgeiiebntigatgvshsaidlgttsakitlraaaghtiyedpitygst 300
QY 301 SYVDALINIPDDGDNKEYTIVFSEGLTEFAKDEKRTSKLQNVAFKNGTVLKG 360
DB 301 syvdalinipddgdnkeytivyfseglteafakdekrtsklqnvafkngtvllkg 360
QY 361 DVLVLSANGFSQDANSKSLIMDGTSLVNTESIELTNLEINIDSLRNCKIKLSAATAQKD 420
DB 361 dvlvlsangfsqdanskslimdgtslvntesieltnleindslrnckiklsataqkd 420
QY 421 IRRDRPVVLAISDESFQNGFLNEDHSYDGLIELDACKDIYISADSNINAVSPYICG 480
DB 421 irrdrpvvlaidesefqngflnedhsydglieldackdiyisadsninavspyicg 480
QY 481 KMTINMGTDKKAATVSAKOSFNPTAEOEAPLVNPLMGSEFIDVRPONTLELTGEGAPY 540
DB 481 kmtinmgtdkkaatvsaakosfnptaeeaplvnplmgsefidvrpontleltgegapy 540
QY 541 EKRFWAGISNVLHRSQRENOQRFRHVSAGAVGASTRMGCDTSLGFAQLFARDQYF 600
DB 541 ekrfwagisnvlhrsqrenqrfrihvsagavvgastrmgcdtslgfaqlfardqyf 600
QY 601 MNINFAKTYAGSLRLQHDASLYSVSLGEGIRELLEPYVSKTLPCCSTYGGOLSTGHTD 660
DB 601 mninfaqtyagslrlqhdaslysvslleggirellepyvsktlpcstygqolsyghtd 660
QY 661 HRKKTSLPPPPPLTSDHRSWGGYVWAGELGRVAVENTSGGEFFERYPPFKVQOAVYS 720
DB 661 hrkktslpppppltsdhrrswggyywagelgrvaventsggefferyppfkvqovays 720
QY 721 RQDSFVELGAISRDFSDSHLYNLAIPLGILKLEKRFEAQYHYVAVMSPDYCRSNPKTYY 780
DB 721 rqsdfvelgaistrdfsdshlynlaiplgilklerfeyahyvavmspdycrsnpkctt 780
QY 781 LLSNGQSWTKGSLNARQAGTIVASGFRSLGAAAELEFGNFGFEMRGSSRSYNAVDAKSK 840
DB 781 llsngqswtkgslnarqagtivagsgfrslgaaaelefgnfgrgssrszynavdagsk 840
QY 841 F 841
DB 841 f 841

```

# RESULT 3

AAW88431

ID AAW88431 standard; Protein; 610 AA.

XX AAW88431;

AC AAW88431;

XX 26-APR-1999 (first entry)

XX

DE Chlamydia pneumoniae surface exposed protein Omp7.

XX Omp7; outer membrane protein 7; surface exposed protein; antigen;

XX Infection; diagnosis; vaccine; atherosclerosis; asthma.

KM

XX Chlamydia pneumoniae.

OS

XX W09586953-A2.

PN

XX 30-DEC-1998.

PD

XX 19-JUN-1998; 98MO-DK00266.

PE

XX 23-JUN-1997; 97DK-0000744.

PR

XX (BIRK/) BIRKELUND S.

PA (CHR1/) CHRISTIANSEN G.

XX Birkelund S, Christiansen G, Knudsen K, Madsen A;

PI Mylind P;

XX WPI; 1999-105610/09.

DR N-PSDB; AAX06830.

XX

PT Species-specific test for identifying mammals infected with

PT Chlamydia pneumoniae - comprises detecting antibodies specific for

PT outer membrane proteins of C. pneumoniae or nucleic acids encoding

PT these proteins

XX

XX Disclosure; Page 86-88; 115pp; English.

XX

CC This polypeptide is described as a subsequence of the claimed

CC novel surface exposed protein Omp7 (see AAW88420) of Chlamydia

CC pneumoniae, a human respiratory pathogen. The invention provides

CC a new species specific test for identifying mammals (including

CC humans) infected with C. pneumoniae. The test comprises detecting

CC antibodies specific for surface exposed proteins Omp4-Omp5 (see

CC AAW88417-28) or detecting nucleic acid fragments encoding them (see

CC AAX06816-27), especially by PCR. The proteins are also used in the

CC diagnosis of C. pneumoniae infection in mammals. The nucleic acids

CC and proteins can also be used in the immunization of mammals, the

CC nucleic acids being particularly useful as DNA vaccines for

CC effecting in vivo expression of antigens. The vaccines may also

CC prevent atherosclerosis and bronchial asthma, which are possibly

CC associated with C. pneumoniae.

XX

XX

SQ Sequence 610 AA:

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 60.5%; Score 2600; DB 20; Length 610;

Best Local Similarity 71.5%; Pred. No. 3.5e-167;

Matches 532; Conservative 21; Mismatches 51; Indels 140; Gaps 8;

```

QY 30 LLSANSPFDGTTSTTSFSK-TSSATDGTNVVFKDSVYIENVPTGQTSCKFNDAAG 88
DB 4 lgsrdyngdtstetefpkatsdaagtyllgdvsiagqyq-tsllscsn--tag 60
QY 89 DLNPLGGFSPTSNIDATTAGAAIGSEAAKNTVTLSCFSALSFLKSPASTVNTNGAI 148
DB 61 nlflfngfslhfdnlstsvagvvnstaaagtkfsgfstlrmaaplt----- 112
QY 149 NVKGNLSLDNDKVLIDNFPSTGDCGAINCAGSLKIANKSLFIGNSSSTRGGAITHKN 208
DB 113 ----- 112

```

outer membrane proteins of *C. pneumoniae* or nucleic acids encoding these proteins

Claim 7, Page 50-52; 115pp; English.

This polypeptide comprises the novel 89.7 kDa surface exposed protein Omp7 of the human respiratory pathogen *Chlamydia pneumoniae*. Its amino acid sequence was deduced from DNA (see AAX06819) isolated from a *C. pneumoniae* expression library. The invention provides 12 novel surface exposed proteins, Omp4-Omp15 (see AAX06816-27), and nucleic acid sequences encoding them (see AAX06816-27). A new species specific test is claimed that is used to identify mammals (including humans) infected with *Chlamydia pneumoniae*. The test comprises detecting antibodies specific for Omp4-Omp15 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used in the diagnosis of *C. pneumoniae* infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with *C. pneumoniae*.

Sequence 841 AA:

Query Match 100.0%; Score 4298; DB 20; Length 841;  
Best Local Similarity 100.0%; Pred. No. 1.9e-281;  
Matches 841; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKIPRLFLISLVPTLSMSNLGAATTEELISANSFDDGTTSTTSFSSKTSATDGTNYVF 60  
1 mkiprlflislvptlsmsnllgaatteelsansfdgtsttsfssksatdgtnyvf 60  
61 KRSVYIENPKTGEFQVSTSCFNDAAGDLNPLGCGFSTESNIDATTSAGAIISSEAN 120  
61 krsvyienpkytgefstscfnkdaagdlnlpgcgfstesnidatlsagaissean 120  
61 KDSVYIENPKTGEFQVSTSCFNDAAGDLNPLGCGFSTESNIDATTSAGAIISSEAN 120  
61 kdsvyienpkytgefstscfnkdaagdlnlpgcgfstesnidatlsagaissean 120  
121 KVTWTSGFSALEFLKSPASTVNGIGAIWVKNLSDLDYKLIQDNFTSGDGGAINCAG 180  
121 kvtwtsgfalsflkspastvngigaiwvknlsldldyqliqdnftsgdggaincag 180  
121 KVTWTSGFSALEFLKSPASTVNGIGAIWVKNLSDLDYKLIQDNFTSGDGGAINCAG 180  
121 kvtwtsgfalsflkspastvngigaiwvknlsldldyqliqdnftsgdggaincag 180  
181 SLKIANKSLSEFIGNSSSTRGAIHTKMLTLLSSGGETLFQGNATPAAGKGAIALDSG 240  
181 slkiankslsefignssstrgaihtkmltllssggetlfqgnatpaagkgaialsdg 240  
181 slkiankslsefignssstrgaihtkmltllssggetlfqgnatpaagkgaialsdg 240  
241 TTSISGSDSDITFEENGTATGTVSHSAIDLGTSAKITALRAQGHITTYFYPTVYTGST 300  
241 ttsisgdsdgitfeengtatgtvshsaildgtlsakitalraqghitlyfyptvtgst 300  
301 SVADALNTNSPPTGNKEVGTIVTSGEKLTEAKDEKNRTSKLONVAFNGTMYLKG 360  
301 svadalntnspptgnkevtgtivtsgeklteakdeknrtskllonvafngtmylkg 360  
301 svadalntnspptgnkevtgtivtsgeklteakdeknrtskllonvafngtmylkg 360  
361 DVLVSANGFSQDANSKLLMDLGTSLVANTESIELTNLEINIDSLRNGKKIKLSAATAOKD 420  
361 dvlvsangfsqdansklldlgtslvantesieltlnleinidslrngkkiklsataokd 420  
361 dvlvsangfsqdansklldlgtslvantesieltlnleinidslrngkkiklsataokd 420  
421 ITIDRPVYLAISDESEFYQNGFLNEDHSYDGLIELDAGKDIVISAOSRSINAVQSPRYG 480  
421 itidrpvylaisdesefyqngflnedhsydglieldagkdivisaosrsinavqspgyg 480  
421 itidrpvylaisdesefyqngflnedhsydglieldagkdivisaosrsinavqspgyg 480  
481 KWTIMNSTDDKRAATVMAKQSFNPAEAPLVPNLLMSGFIDVDFPQFIELEGAGY 540  
481 kwtimnstddkrtatvmaqsfnptaepaplvpnlmsgfidvdfpqfielelegagy 540  
481 kwtimnstddkrtatvmaqsfnptaepaplvpnlmsgfidvdfpqfielelegagy 540  
541 EKRFFWAGISNVLNHSRGRENQKRFHVSGCAVGASTRMPGSDDTLSLGAQLFARDKYF 600  
541 ekrffwagisnvlhnsrgrenqkrfhvsagavgastrmpgddtllslgaqlfardkyf 600  
541 ekrffwagisnvlhnsrgrenqkrfhvsagavgastrmpgddtllslgaqlfardkyf 600  
601 MNTNFAKTYAGSLRIQHDLSLVSYSIILGEGGLREIILPYVSKTLPCEFYQOLSTGHND 660  
601 mntnfaakyagslriqhdslysvsillgegglreillpyvsktlpcsfyqolsyghnd 660  
601 mntnfaakyagslriqhdslysvsillgegglreillpyvsktlpcsfyqolsyghnd 660

661 HRMKTESLPPPEPTLSTDBTSGWYVWAGELCTRAVENTSGRGEFFREYTPVKVQAVYS 720  
661 hrmkteslpppplstdbtsgwyvwagelctraventsgrgffreytpltvkqvays 720  
721 RODSEVELGAISSRDSHLYMIAIPLGIKLEKREAEQYHYVAVMSPPDVCSNPKCTTT 780  
721 rodsevelgaissrdsdhlymiaiplgiklekreaeqyhyvavmsppdvcsnpkcttt 780  
781 LLSNGSKMTKSSNLAHQAGIYQASGFRSLGAAELFNGFGEEMGSSSYNDAGSKIK 840  
781 llsngskmtkssnlaragiyqasgfrslgaaelfngfgeemgsssyndagskik 840  
841 F 841  
841 f 841

# RESULT 2

ID AAY92818 standard; Protein; 841 AA.

AY92818;

29-AUG-2000 (first entry)

C. pneumoniae CPN100397 antigen.

Antigen: anti-inflammatory; respiratory; antibacterial; anti-asthmatic;

anti-arteriosclerotic; vaccine.

*Chlamydia pneumoniae*.

WO200024765-A2.

04-MAY-2000.

28-OCT-1999; 99WO-CA00992.

28-OCT-1998; 98US-0106034.

28-OCT-1998; 98US-0106039.

28-OCT-1998; 98US-0106042.

28-OCT-1998; 98US-0106044.

29-OCT-1998; 98US-0106072.

29-OCT-1998; 98US-0106073.

29-OCT-1998; 98US-0106074.

29-OCT-1998; 98US-0106087.

02-NOV-1998; 98US-0106587.

02-NOV-1998; 98US-0106588.

02-NOV-1998; 98US-0107034.

02-NOV-1998; 98US-0107035.

(CONN-) CONNAUGHT LAB LTD.

Muridin AD, Comen RP, Wang J;

WPI: 2000-350688/30.

N-PSDB: AAA28690, AAA28691.

Chlamydia antigens and the proteins they encode, useful for

vaccinating against *Chlamydia* infections that affect the respiratory tract

Claim 13; Fig 1; 226pp; English.

The nucleic acids may be used for the recombinant production of the *Chlamydia* polypeptides (either in vivo or in vitro) according to standard recombinant DNA methodologies. The polypeptides may then be used to vaccinate against *Chlamydia* infections in mammals. *Chlamydia*, such as *C. pneumoniae*, are pathogens responsible for upper respiratory tract infections such as community acquired pneumonia, acute respiratory disease and bronchitis and may be implicated in atherosclerotic changes and asthma. The nucleic acids may also be used as probes for detecting

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2001, 03:26:29 ; Search time 97.54 Seconds  
(Without alignments)  
522.706 Million cell updates/sec

Title: US-09-446-677B-8  
Perfect score: 4298  
Sequence: 1 MKIPRLFLISLVPTLSMSN.....FEWGRSSRYNDAGSKIKF 841

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_0601:\*

- 1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4298	100.0	841	20	AAW88420	Chlamydia pneumoniae
2	4271	99.4	841	21	AA192818	C. pneumoniae CPN1
3	2600	60.5	610	20	AAW88431	Chlamydia pneumoniae
4	2499	58.1	483	20	AA134609	Chlamydia pneumoniae
5	1564	36.4	928	21	AA194327	Chlamydia pneumoniae
6	1553	36.1	928	20	AAW88421	Chlamydia pneumoniae
7	1429.5	33.3	928	20	AAW88417	Chlamydia pneumoniae
8	1429.5	33.3	928	21	AA190236	Chlamydia antigen
9	1429.5	33.3	949	20	AA135060	Chlamydia pneumoniae
10	1413.5	32.9	930	20	AA135052	Chlamydia pneumoniae
11	1409.5	32.8	918	21	AA169369	Amino acid sequence

12	1407.5	32.7	930	21	AA190240	Chlamydia antigen
13	1404.5	32.7	930	20	AAW88424	Chlamydia pneumoni
14	1400.5	32.6	928	20	AAW88418	Chlamydia pneumoni
15	1394.5	32.4	918	20	AAW88422	Chlamydia pneumoni
16	1374.5	32.0	928	21	AAW88423	Chlamydia pneumoni
17	1369.5	31.9	928	21	AAW88429	Chlamydia antigen
18	1364.5	31.7	928	21	AA190237	Chlamydia antigen
19	1346.5	31.3	914	20	AAW88429	Chlamydia pneumoni
20	1330	30.9	927	20	AA135054	Chlamydia pneumoni
21	1329.5	30.9	926	21	AA199842	Chlamydia pneumoni
22	1315.5	30.6	925	21	AA199843	Chlamydia pneumoni
23	1308	30.4	885	21	AA199238	Mature Chlamydia a
24	1186.5	27.6	945	21	AA169368	Amino acid sequenc
25	1164.5	27.1	945	20	AAW88428	Chlamydia pneumoni
26	1144	26.6	643	20	AA135056	Chlamydia pneumoni
27	1084	25.2	973	21	AA196274	Chlamydia POMP91B
28	1000	23.3	922	21	AA195548	Chlamydia pneumoni
29	999	23.2	922	20	AA134597	Chlamydia pneumoni
30	996	22.3	922	20	AAW88419	Chlamydia pneumoni
31	957.5	22.3	597	20	AA134611	Chlamydia pneumoni
32	937	21.8	1006	21	AA136339	C. trachomatis pmp
33	932.5	21.7	1012	20	AA116735	C. trachomatis LGV
34	931	21.7	1013	20	AA116737	C. trachomatis B s
35	929	21.6	982	21	AA136333	C. trachomatis pmp
36	928	21.6	1013	20	AA116738	C. trachomatis F s
37	920	21.4	199	20	AA134608	Chlamydia pneumoni
38	861	20.0	202	20	AA134607	Chlamydia pneumoni
39	837.5	19.5	671	20	AA135050	Chlamydia pneumoni
40	795	18.5	1132	20	AA135048	Chlamydia pneumoni
41	788	18.3	507	20	AA134614	Chlamydia pneumoni
42	728.5	16.9	530	20	AA135064	Chlamydia pneumoni
43	708.5	16.5	880	21	AA113632	C. trachomatis pmp
44	695.5	16.2	866	21	AA113638	C. trachomatis pmp
45	632.5	14.7	450	20	AA134617	Chlamydia pneumoni

#### ALIGNMENTS

RESULT 1	
AAW88420	standard; Protein; 841 AA.
XX	
XX	AAW88420:
AC	
DT	26-APR-1999 (first entry)
XX	
DE	Chlamydia pneumoniae surface exposed protein Omp7.
XX	
KM	Omp7: outer membrane protein 7; surface exposed protein; antigen;
KW	infection; diagnosis; vaccine; atherosclerosis; asthma.
OS	Chlamydia pneumoniae.
XX	
XX	MO9858953-A2.
PN	
XX	
PD	30-DEC-1998.
XX	
PF	19-JUN-1998; 98WO-DK00266.
XX	
PR	23-JUN-1997; 97DK-0000744.
XX	
PA	(BIRK/) BIRKELUND S.
PA	(CHR1/) CHRISTIANSEN G.
PI	Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI	Mygind P;
XX	
DR	WPI, 1999-105610/09.
DR	N-PSDB; AAX06819.
XX	
PT	Species-specific test for identifying mammals infected with
PT	Chlamydia pneumoniae - comprises detecting antibodies specific for

**THIS PAGE BLANK (USPTO)**

BEST AVAILABLE COPY

---

Db 541 GACGTTACCATCTTAAGTAATGATATCTATACCCCTAATTTTCTAAAGAAA 596

# RESULT 15

AM085457

LOCUS 451 bp mRNA EST 14-OCT-1999

DEFINITION wy65d10.x1 Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1 Homo sapiens cDNA clone

IMAGE:2553427 3', mRNA sequence.

ACCESSION AM085457

VERSION AM085457.1 GI:6040609

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 451)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

This clone is available royalty-free through LIND ; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Seq primer: -40UP from GIDCO

High quality sequence stop: 363.

Location/Qualifiers

1. 451

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2553427"

/clone\_1ib="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT7r3D-Pac (Pharmacia) with

a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

Equal amounts of plasmid DNA from five normalized

libraries were mixed, and ss circles were made in vitro.

Following HAP purification, this DNA was used as tracer in

a subtractive hybridization reaction. The driver was

PCR-amplified cDNAs from pools of 5,000 clones made from

the same 5 libraries. The pools consisted of the following

libraries and clones: Soares NBHSF pool 1:

309384-310919, 323208-325895 Soares NB2HP pool 1:

145032-147335, 147720-148103, 148872-149255, 15002 -

150407, 151176-152327 Soares NB2HP8-9W pool 1:

758280-760583, 772104-774407 Soares NBHPA pool 1:

304776-306311, 320136-322823, 326280-326663 Soares NBHOF

pool 1: 723720-726407, 739080-740999 Subtraction by Bento

Soares and M. Fatima Bonaldo."

BASE COUNT 136 a 82 c 68 g 165 t

ORIGIN

Query Match 1.5%; Score 39; DB 111; Length 451;

Best Local Similarity 53.6%; Pred. No. 4.9;

Matches 81; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

0y 1128 GTTGATTATGATTAGGAGCTGTTGGTGCACACACCGAAGTATCGATTACGAA 1187

Db 62 GTTATTATGATTATTTTCACACTTCGCCGTTGGGATATCACAAAGGTTCTTGGCCAC 121

0y 1188 TTTGGAAATTAATATAGACTCTCTCAGACGGGAAAAAGATAAACTCAGTGTCCAC 1247

Db 122 GAAACAAACAAACACATTTAAACCAACAAACAGCAACAGCAGTCAGTCTGCAGC 181

0y 1248 AGCTGAGAAAGATATTCGTATGATCGTCT 1278

Db 182 AGCAGAGGAGATTAATGTTGATCATCT 212

Search completed: October 2, 2001, 03:06:46

Job time: 5164 sec

Accession	Version	Keywords	Source	Organism
BACR01J16 of RPCI-98	fly1	genomic survey sequence.		Drosophila melanogaster (fruit fly)
AL065414	AL065414.1	GI:4938827		Drosophila melanogaster
REFERENCE	AUTHORS	TITLE	JOURNAL	
COMMENT	<p>Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mamosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a>.</p>			
FEATURES	<p>Source</p> <p>1. 1101</p> <p>/organism="Drosophila melanogaster"</p> <p>/db_xref="taxon:7227"</p> <p>/clone_lib="RPCI-98"</p> <p>/clone="BACR01J16"</p> <p>/note="end : 1101"</p>			
BASE COUNT	<p>280 a 104 c 123 g 211 t 383 others</p>			
ORIGIN	<p>1.6% Score 39.4; DB 219; Length 1101;</p> <p>Best Local Similarity 16.3%; Pred. No. 5.1;</p> <p>Matches 77; Conservative 183; Mismatches 211; Indels 0; Gaps 0;</p>			
Query Match	<p>1.6% Score 39.4; DB 219; Length 1101;</p> <p>Best Local Similarity 16.3%; Pred. No. 5.1;</p> <p>Matches 77; Conservative 183; Mismatches 211; Indels 0; Gaps 0;</p>			
1173	TATCGAGTTAAGAAATTTGGAATTAATTAAGACTCTCTCGAAGACGGGAAAGATATAA	1232		
567	TWTATTWKAANAAAAAAMAAAAAAMAAAAAAGARGATNDDKBKADKAAADAKAKAA	626		
1233	ACTGAGCTGCTGCCACAGCTCAGAAAGATATTCGTATAGATCGTCTGTTGTACTGGCAAT	1292		
627	WHAGTWTWTKKPTDKKKKKKKKKKKKKRAATDDTATATWTKAKDKAKAAKATADRTK	686		
1293	TAGCAGTAGAGATTTTTATCAAAAGCGCTTTTGAATGAGACCACTCTCTAATGARGGAT	1352		
667	AMKGAGGAGARAKAARAKAPAKKRWMDDDDTKKKKKTTATKTTTTRAAADKKMKWRK	746		
1353	TCTTCAATGAGATGCTGGGAAAGACATGCTGATTTCTCAGATCTCCAGATATAATGC	1412		
747	KATAGAMWDDARRRRGGGWRGMDKRRRTARAGDKDTGKKKTKGDRKDGGGDGDGDA	806		
1413	TGTACAATCTCCGATGCTATCAGAGGAAGTGACCAATCAATGGCTACTGATGATTA	1472		
807	GGGDRDRBARARRAMWDPKDTRGAKDKADKKKAGKRKMWPRKTKRDTADMKAAAMRW	866		
1473	GAAGCTACGGTTCTTGGGCAAGCAAAAGTTTAATCCACACTGCTGAGAGAGAGCTCC	1532		
867	FWRRRRKRRRRGRDRWRWRGADKKWKKDDDDTADKDDWTKKDDTWTAGATGAMDDAKAD	926		
1533	GTTAGTTCCTAATCTCTTGGGGTCTTTTATAGATTTGGTGCCTCCAAATTTAT	1592		
927	GDWKGITWRTWTKRWKMGRRGDDGKRTTAGAAGGKKKKGKKDKDRDDAKATDGGKTTW	986		

QY	1593	AGAGTAGTACTGAAGGTGCTCTACGAAAGAGATTTGGGTGCAG	1643
Db	967	RDYTTAKRRRRRRRRRACRGDRGRGDKKKRRARWRKCKKKTKDDDD	1037
RESULT	14		
LOCUS	BF979823	870 bp	EST
DEFINITION	60228833371 NIH_MGC_97	Human sapiens cDNA clone IMAGE:4374032.3',	
ACCESSION	BF979823	NIH_MGC_97	
VERSION	BF979823.1	GI:12347038	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
TITLE	1 (bases 1 to 870)		
JOURNAL	NIH-MGC http://imgc.ncbi.nlm.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabs-remail.nih.gov		
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.		
	cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki		
	Toshiyuki and Piero Carninci (RIKEN)		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNLN at:		
	http://image.llnl.gov		
	Plate: LHAM10037 row: a column: 09		
	High quality sequence start: 18		
	High quality sequence stop: 740.		
FEATURES	Location/Qualifiers		
Source	1..870		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4374032"		
	/clone_lib="NIH_MGC_97"		
	/lab_host="DH10B"		
	/note="Organ: testis; Vector: pBluescript (modified		
	pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtagag		
	); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',		
	size-selected for average insert size 2.2 kb and		
	normalized to R01 5. This is a primary library enriched		
	for full-length clones and constructed using the		
	Cap-trapper method (Carninci, in preparation). Library		
	constructed by M. Brownstein (NIH/NHGRI, National		
	Institutes of Health). Note: this is a NIH_MGC Library."		
BASE COUNT	298 a 158 c 148 g 266 t		
ORIGIN			
Query Match	1.6%, Score 39.2; DB 172; Length 870;		
Best Local Similarity	47.9%, Pred. No. 5.3;		
Matches	113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;		
QY	751	ATTATCTTTGAAGCAATACGATAGGACATACGAGACCGTCTCATAGTGTATTCAT	810
Db	361	AATATGCTAAAGGTAAAGCATTAAGATACAAAGATCTCACAGAGGTTAATATTTTACA	420
QY	811	TTAGAGACTAGCGCTAAATATACGCTGTTAGCTGGCGAAGGACATACGATATACTT	870
Db	421	ACACTAAAAAATAAATGCTCTATATATTTCTTAGTGGGACATTTGCTTCAATT	480
QY	871	TATGATCCGATTACTGTACACAGATCGACATCTGTGGCTGATGCTCTCAATATTAATAGC	930
Db	481	TAAATTTTGGTTATGCTTAACAAAATATACGCTCTAAAGAGAGTTTCAAAAAAATGCAAT	540
QY	931	CTGTACTCGAGATACAAAGAGATATACGGGACCAATAGTCTTTTCTGGAGAGAA	986



D <sub>b</sub>	406	ACTACTACTGCTGTACTAGTGGTGTAACACKCGCKGTGTTACTACTACTACTACT	347
OY	733	TCTGGACACAGTAGCCGCATATTATCTTTGAAGGCCAATACGATAGGAGCTAACAGCAACCCTC	7922
D <sub>b</sub>	346	ACTCATCTTTTCCTACTGGTACTACTACTACTACGCTACTACTACTACTACTACTACTACT	287
OY	793	TCTCATATGCTGATTGGATTAGAACTAGCGGTCAATAGATTAAGTGGTACGNGCGCAA	8524
D <sub>b</sub>	286	ACTACTACTRRTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	227
OY	853	GGACATAACGATATACCTTTTTATGATCCGATTACTGTAAACAGATGCACATCTGTGGTGAAT	9122
D <sub>b</sub>	226	GCTCTACTACTCACACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	167
OY	913	GCTCTCAATTTATTAACCCCGATACT	939
D <sub>b</sub>	166	GCTCTACTACTACTACTACTACTACTACT	140

RESULT	11			
CNS00F3W/C				
LOCUS	CNS00F3W	1021 bp	DNA	GSS
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BAC30019 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	AL070005			
VERSION	AL070005.1			
KEYWORDS	GI:4950148			
SOURCE	GSS.			
	fruit fly.			

QY	350	AAGACGCTAAACACAGTACCGTTATCAGAGTTTGGCACTCTTTCTTAATCCC	409
Db	745	GAGAGGGTAATAAAATTTGTMA-TTGGTTAAATTGTTGARGKTAAATTTTTTTTTTCCC	687
QY	410	CAGCAAGTACAGTCACCTAATGGATTGGAGCTATCAATGTTAAAGGAATTTAAGCCTAT	469
Db	686	CCCYCYMCHMYCAYMAATTGMAKYCAGATKTTTWTGTCMAATGTGGGGGTCTGCTCT	627
QY	470	TGCTAATGATAGGCTTTTGATTCAGACAATTTCTCAACAGAGATGGCGGACAAATTA	529
Db	626	CAGTTTCAACATCTCGCCACATAAATATYTTTCATCAGGGGGGGAAGGAAACMAAGTTC	567
QY	530	ATTGTGCAG	538
Db	566	MGAGTGCAG	558

RESULT	12
LOCUS	AV675196
DEFINITION	AV675196 573 bp mRNA EST 05-OCT-2000 CDNA clone c1b1113 5', mRNA sequence.
ACCESSION	AV675196
VERSION	AV675196.1 GI:10113195
KEYWORDS	EST.
SOURCE	Ciona intestinalis.
ORGANISM	Ciona intestinalis. Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia; Clonidae; Ciona. 1 (bases 1 to 573) Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T. Expressed genes in Ciona intestinalis Unpublished (2000) Contact: Nori Satoh
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

## FEATURES

### Location/Qualifiers

Query Match	1.6%;	Score 39.6;	DB 32;	Length 573;
Best Local Similarity	51.7%;	Pred. No. 3.6;		
Matches 90; Conservative	0;	Mismatches 84;	Indels 0;	Gaps 0

QY	1026	TTCTAAATTTCTTCCAAAATGTCCTTTTAAANAATGGAGCTGTAGTGTTTTAAAGGTGATGT	1085
Db	354	TTATTAAGAATCTGGAGAAAGAAAGAACCCCTATGTGGCGAATACGTAAAGAAATTTGT	413
QY	1086	CGTTTTAAGTGCAGACGGTTCTCTCAGATGCAGAACCTAAGTTGATGTATGATTTAGG	1145
Db	414	TATTTTGAAGCGCAAAAGGTTCTTATGGAGATGCAATTTAAATTTAGTGCCTATTTCACA	473
QY	1146	GACGTGCTGGTTGCAACACCGCAAGATGTACAGATTTACGAATTTTGGAAATTTAA	1199
Db	474	AACTTTATGCGCGATCATGATGACATGTTGGAGTTAGCAGAACTCTCATTTAA	527

RESULT	13			
CNS000D1				
LOCUS	1101 bp	DNA	GSS	03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC #			

	Query Match	1.6%;	Score 40.6;	DB 119;	Length 580;
	Best Local Similarity	51.4%;	Pred. No. 1.8;		
	Matches 94;	Conservative 0;	Mismatches 89;	Indels 0;	Gaps 0
Qy	23	CTAGTGTCTTTAAATGACGCTGCAGCGTCAATCTAAATTTCTTAGGGGGGATTTT	292		
Db	28	CTGTCTCTTTTAAATGATGATGAAAGATGGAGAAATCTGAATCTAAAGAACAAGAACTC	87		
Qy	293	CTTTCACATTTTGCATATTCGATGCAACACAGGCTCTGAGCGTGTATTTGGAAGTGAAG	352		

Query Match	1.6%	Score 39.8	DB 221	Length 735
Best Local Similarity	42.4%	Pred No. 3.4	Mismatches 215	Indels 0
Matches 164	Conservative	8	Mismatches 215	Indels 0
QY	553	GCAAACAAATAGTCCTCTTTATTTGAAATAGTCTTCAACAGTGGCGAGGATT	612	
Db	526	GCTAAACACTACTAGCTACTACTGAGCTGCTACTACTACTAGCTACMATTTCMGCTGCT	467	
QY	613	CATACCAAAAACCTACACTATCTCTGTGGTGGGAACTCTATTAGGGAATTAAGCG	672	
Db	466	ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	407	
QY	673	CCTACGGCTGCTGGTAAAGAGGTGATCGGATTCAGACTCTGGCAACCCATTCATT	732	

source 1. 778  
 /organism="Homo sapiens"  
 /db.xref="taxon:9606"  
 /clone="CLOB0032A04"  
 /clone\_lib="LTI\_NFL006\_PU2"  
 /issue\_type="Placenta"  
 /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end clones, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 138 a 169 c 133 g 290 t 48 others  
 ORIGIN

Query Match 1.7%; Score 43.4; DB 105; Length 778;  
 Best Local Similarity 44.9%; Pred. No. 0.31;  
 Matches 93; Conservative 16; Mismatches 98; Indels 0; Gaps 0;

OY 888 AACAGATGACATCTGCTGATGCTCTCAATATTATACCCCTGATACGAGATGA 947  
 : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : ||||  
 DB 269 RAAGAGAAARARCGTGTGTGTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 210  
 OY 948 CAAAGATGATGAGGACATATGCTTTCTGAGAGAAAGCTCAGGAGGAGAACTGA 1007  
 : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : ||||  
 DB 209 GGCAGAGCTACCAAAACAGCCCTGAAARAAAGCTGAAAGAAAGAAAGCTA 150  
 OY 1008 AATGAGAGAGACCCGACTTAATTAATTAATTAATTAATTAATTAATTAATTA 1067  
 : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : ||||  
 DB 149 AAGAGGAAARAAAGCTGTGCGCAAAAATTTAATTGATTTTCCAAAAGCTAA 90  
 OY 1068 AGTTTAAAGGTGATGCTTTTAAG 1094  
 : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : ||||  
 DB 89 AATTGTAAGCTTTTGAAMAAR 63

RESULT 7  
 BE187133/c 450 bp mRNA EST 22-JUN-2000  
 LOCUS  
 DEFINITION NXNV\_159.D05\_F Nsf Xylem Normal wood Vertical Pinus taeda cDNA  
 accession BE187133  
 version BE187133.1 GI:8666317  
 keywords  
 source  
 organism Pinus taeda  
 loblolly pine.

REFERENCE 1 (bases 1 to 450)  
 Sederoff, R.  
 Molecular Basis of Wood Formation in the Pine Megagenome  
 Unpublished (2000)  
 contact: Johnson, Arthur  
 North Carolina State University  
 Tel: 919 515 7800  
 Fax: 919 515 7801  
 Email: ajohnson@unity.ncsu.edu  
 Seq primer: T3.

FEATURES  
 source  
 location/Qualifiers

1..450  
 /organism="Pinus taeda"  
 /db.xref="taxon:3352"  
 /clone="NXNV\_159.D05"  
 /clone\_lib="Nsf Xylem Normal wood Vertical"  
 /note="Vector: Bluescript SK; Site 1: Eco RI; The sequences contain a 'cDNA adapter' between the EcoRI site and the start of the EST. The adapter sequence is 'AATTCGCGACGAG'."

BASE COUNT 146 a 60 c 105 g 124 t 15 others  
 ORIGIN

Query Match 1.7%; Score 42.4; DB 164; Length 450;  
 Best Local Similarity 52.2%; Pred. No. 0.5;  
 Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

OY 130 ACCTTTTCAGAAACATCATCGGTACAGATGGACCAATTTAGTTTAAAGATTCT 189  
 : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : ||||  
 DB 398 AACNNATAGTAGCAAAAGACANCTGTGCTGTGACCAAGATTTTAAATTAATCT 339  
 OY 190 GTAGTTATGAAATGTACCAAAACAGGGAACCTACTACTAGTTGTTTAAAT 249  
 : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : ||||  
 DB 338 TTAATTTGCAACCATCAATCAATTAATTAATTAATTAATTAATTAATTAATTA 279  
 OY 250 GACGCTGCAGCTGAGATCTAATTTCTTAGAGAGGG 286  
 : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : ||||  
 DB 278 TCCATTTGCTGTTGNCCTTCATATATGTAACCAATGG 242

RESULT 8  
 CNS017KX/c 1101 bp DNA GSS 26-JUL-1999  
 LOCUS  
 DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC  
 BACN37F10 of DrosBAC library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 accession AL108171  
 version AL108171.1 GI:5628475  
 keywords  
 source fruit fly.

ORGANISM  
 Plasmid Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 1101)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Direct Submission  
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the European Drosophila Genome Project (EDGP) -  
 http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
 d'Etude du Polymorphisme Humain) with funding provided by a MRC  
 project grant. The DNA was prepared from embryos by Alain Bucheton  
 and Genevieve Payan. It has been constructed in the vector  
 pBelobAC11.

FEATURES  
 source  
 location/Qualifiers

1..1101  
 /organism="Drosophila melanogaster"  
 /plasmid="pBelobAC11"  
 /db.xref="taxon:7227"  
 /clone\_lib="DrosBAC"  
 /clone="BACN37F10"  
 /note="end : SP6"

BASE COUNT 247 a 84 c 159 g 263 t 348 others  
 ORIGIN

Query Match 1.6%; Score 41.4; DB 219; Length 1101;  
 Best Local Similarity 24.3%; Pred. No. 1.3;  
 Matches 64; Conservative 87; Mismatches 112; Indels 0; Gaps 0;

OY 1306 TTTTATCAAAATGCTTTTGAATGAGACACATTCATGATGATGATGATGATGAT 1365  
 : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : ||||  
 DB 1022 TTAATDPTWKRTTGTGTRTKWMAARRAKWRDWDMDKMDTKDKDRMTKRWAT 963  
 OY 1366 GCTGGGAAGACATGCTATTTCTGACAGATTCGTCAGATTAAGCTGACAACTCGC 1425  
 : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : ||||  
 DB 962 RGDGAGRAKRRDRDGRGRRRRRRRTKRWRRRAAMWTATWTWGTGKAWMTYTR 903

[illegible]

**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - FRANCE (E-mail : segrete@genoscope.cns.fr)  
**COMMENT** Web : www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org> The BDGP Drosophila  
 melanogaster BAC library was prepared by Kazutoyo Oseguwa and  
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCI-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
 P1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
**FEATURES**  
 source Location/Qualifiers  
 1..1178  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="RPCI-98"  
 /clone="BACR12E18"  
 /note="end : TET3"  
**BASE COUNT** 226 a 151 c 164 g 438 t 199 others  
**ORIGIN**  
 Query Match 1.8%; Score 44.4; DB 219; Length 1178;  
 Best Local Similarity 36.4%; Pred. No. 0.18;  
 Matches 86; Conservative 42; Mismatches 108; Indels 0; Gaps 0.  
 OY 1006 AAGAGTGAACAACCGCACTCTTAATTAATCTCAAAAGTGGCTTTAAAGGAGACT 1065  
 ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
 DB 1174 VMAAAAAAAAAACCCSCAAAMAMKKKTYTTAAATGCTGTTTAAKGGRRAA 1115  
 OY 1066 GTAGTTTAAAAGTGATGTCGTTTAAAGTGGAAACGGTTTCTCAGAGTCAAACTCT 1125  
 ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
 DB 1114 AAATKTTTGAGAAAAAAMAMAMAMAMATTAAMGSGTGTTRAAAMAGRAWTMG 1055  
 OY 1126 AAGTTGATTAGATTAGAGCAGTCGTTGGTTCGACAAACCGAAAGTAGTCAGTTAACG 1185  
 ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
 DB 1054 AGRGKGTCTTKTGCGTGGWARRAARGKGTWAGAGGTCCVAGAAAMATWTDDAAT 995  
 OY 1186 AATTGGAAATTAATATAGACTCTCGACGAGGGAAMAAAGATAAATCACTGATC 1241  
 ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
 DB 994 AMTTTTTATTAATGCGTGKTTTGGGTAAHMAVMAAAATMAAAAAAARAC 939  
**RESULT** 6  
 AL514531/c mRNA EST 13-FEB-2001  
**LOCUS** AL514531 778 bp  
**DEFINITION** AL514531 LTI\_NF1006\_PL2 Homo sapiens cDNA clone CL0BB0032A04 3  
 prime, mRNA sequence.  
**ACCESSION** AL514531  
**VERSION** AL514531.1 GI:12778025  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
**REFERENCE** 1 (bases 1 to 778)  
**AUTHORS** Li W.B., Gruber C., Jessee J. and Polayes P.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** unpublished (2001)  
**COMMENT** Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: segrete@genoscope.cns.fr, Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr).  
 Location/Qualifiers





117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est110:\*  
191: gb\_est111:\*  
192: gb\_est112:\*  
193: gb\_est113:\*  
194: gb\_est114:\*  
195: gb\_est115:\*  
196: gb\_est116:\*  
197: gb\_est117:\*  
198: gb\_est118:\*  
199: gb\_est119:\*  
200: gb\_est120:\*  
201: gb\_est121:\*  
202: gb\_est122:\*  
203: gb\_est123:\*  
204: gb\_est124:\*  
205: gb\_est125:\*  
206: gb\_est126:\*  
207: gb\_est127:\*  
208: gb\_est128:\*  
209: gb\_est129:\*  
210: gb\_est130:\*  
211: gb\_est131:\*  
212: gb\_est132:\*  
213: gb\_est133:\*  
214: gb\_est134:\*  
215: gb\_est135:\*  
216: gb\_est136:\*  
217: gb\_est137:\*  
218: gb\_est138:\*  
219: gb\_est139:\*  
220: gb\_est140:\*  
221: gb\_est141:\*  
222: gb\_est142:\*  
223: gb\_est143:\*  
224: gb\_est144:\*  
225: gb\_est145:\*  
226: gb\_est146:\*  
227: gb\_est147:\*  
228: gb\_est148:\*  
229: gb\_est149:\*  
230: gb\_est150:\*  
231: gb\_est151:\*  
232: gb\_est152:\*  
233: gb\_est153:\*  
234: gb\_est154:\*  
235: gb\_est155:\*  
236: gb\_est156:\*  
237: gb\_est157:\*  
238: gb\_est158:\*  
239: gb\_est159:\*  
240: gb\_est160:\*  
241: gb\_est161:\*  
242: gb\_est162:\*  
243: gb\_est163:\*  
244: gb\_est164:\*  
245: gb\_est165:\*  
246: gb\_est166:\*  
247: gb\_est167:\*  
248: gb\_est168:\*  
249: gb\_est169:\*  
250: gb\_est170:\*  
251: gb\_est171:\*  
252: gb\_est172:\*  
253: gb\_est173:\*  
254: gb\_est174:\*  
255: gb\_est175:\*  
256: gb\_est176:\*  
257: gb\_est177:\*  
258: gb\_est178:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 03:06:34 ; Search time 10300.6 Seconds  
(without alignments)  
2318.115 Million cell updates/sec

Title: US-09-446-677B-7  
Perfect score: 2526  
Sequence: 1 ATGAGATTCCACATCCGCTT.....GTACCAAAATCAATTTTAG 2526

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*
```



Db 3124 TCATCGAGCATCTTTTCTAGTGTCTAGTAGCTGTGATGGAAGCTCGGACCCCTATT 3183  
QY 2086 GCTGTGAAAATACACGCGCAGAGA 2112  
Db 3184 GCTTAGAAAATACGACGATATGAGCA 3210  
RESULT 13  
5196523-5  
; Patent No. 5196523  
; APPLICANT: LEE, AMY S.  
; TITLE OF INVENTION: CONTROL OF GENE EXPRESSION BY GLUCOSE,  
; CALCIUM AND TEMPERATURE  
; NUMBER OF SEQUENCES: 28  
; CURRENT APPLICATION DATA:  
; FILING DATE: 19-MAY-1989  
; FILING DATE: 19-MAY-1989  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 282,880  
; FILING DATE: 05-DEC-1988  
; APPLICATION NUMBER: 690,951  
; FILING DATE: 01-JAN-1985  
; SEQ ID NO:5:  
; LENGTH: 5470  
5196523-5

Query Match 1.3%; Score 32.6; DB 6; Length 5470;  
Best Local Similarity 60.9%; Pred. No. 13;  
Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 2026 TCACGAGCATCTTTCTTGGGAGGATATGCTGGGCTGGAGAGCTGGGAATCGAGTT 2085  
Db 3124 tcctcgagcatcttctctagtcagtcagcttgatggaagcgctggtaccctatt 3183  
QY 2086 GCTGTGAAAATACGCGCAGAGA 2112  
Db 3184 gcttagaaaataccagacagatatgagca 3210

RESULT 14  
US-08-821-994-86/C  
; Sequence 86, Application US/08821994A  
; Patent No. 6228643  
; GENERAL INFORMATION:  
; APPLICANT: Greenland, Andrew J  
; APPLICANT: Thomas, Didier RP  
; APPLICANT: Jepson, Ian  
; TITLE OF INVENTION: Promoters  
; FILE REFERENCE: PPD 50108  
; CURRENT APPLICATION NUMBER: US/08/821,994A  
; CURRENT FILING DATE: 1997-03-22  
; EARLIER APPLICATION NUMBER: PCT/GB97/00729  
; EARLIER FILING DATE: 1997-03-18  
; EARLIER APPLICATION NUMBER: GB 9606062.9  
; EARLIER FILING DATE: 1996-03-22  
; NUMBER OF SEQ ID NOS: 89  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 86  
; LENGTH: 1102  
; TYPE: DNA  
; ORGANISM: Brassica napus  
US-08-821-994-86

Query Match 1.3%; Score 32.4; DB 4; Length 1102;  
Best Local Similarity 47.5%; Pred. No. 6.2; Indels 0; Gaps 0;  
Matches 96; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 1357 GAGTTAGATCTGGGAAAGCATCGTATTCGACAGATTCGCACTATTAATGCTGTA 1416  
Db 550 GGTGTTAATCGCGCGTTTTCATTATGAAATGGAACCATATCATGAGCCCGCTTA 491

QY 1417 CAATTCGCGATATGCGTATTCAGGGAAGTGCATCAATTGCTCTACTGATAGAA 1476  
Db 490 CAGCCTTGTTGATGATGTTGTCGACAGTCGACAAAGTTCTTGTGGACAAAGATCAGT 431  
QY 1477 GCTACGGTTTCTTGGGAAAGCAAGTTTAAATCCACTGCTGAGCAGAGGCTCGTTA 1536  
Db 430 TCTCCTTTAGATCTGTTTATATCTTACTGCTGACGCTGTGAAGCCCAACA 371  
QY 1537 GTTCTTAATCTCTTGGGTT 1558  
Db 370 CTTCGCAAGTTCCTGTCCT 349

RESULT 15  
US-08-357-264-2  
; Sequence 2, Application US/08357264  
; Patent No. 5541077  
; GENERAL INFORMATION:  
; APPLICANT: BURNIE MR., James P.  
; APPLICANT: MATTHEWS MS., Ruth C.  
; TITLE OF INVENTION: FUNGAL STRESS PROTEINS  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cushman, Darby & Cushman  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/357,264  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/152669  
; FILING DATE: 16-NOV-1993  
; APPLICATION NUMBER: US 663897  
; FILING DATE: 14-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KOKULIS, Paul N.  
; REGISTRATION NUMBER: 16773  
; REFERENCE/DOCKET NUMBER: PNK/5544/202253/DJP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-861-3000  
; TELEFAX: 202-822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-357-264-2

Query Match 1.3%; Score 32.4; DB 1; Length 1200;  
Best Local Similarity 44.7%; Pred. No. 6.5; Indels 156; Gaps 0;  
Matches 126; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 1030 AATTAATCTCAAAATGTTGTTTAAAAATGGGACTGATTTTAAAGGTGATGCTT 1089  
Db 493 AATCCCGTGAATAATACCATTTCTTGATGCCCTTGAAGCTTAAGAGCTTGG 552  
QY 1090 TTAAGTCGAAGGTTTCTCTCAGATGCAAACTTAAGTTGATTATGATTAGGAGC 1149  
Db 553 TTCTATGATGATCAATCATGATATGATGATGATGATGATGATGATGATGATGATG 612



Db	2998	TAGTGCAAATTAATTCATTATCTCTTTACATTAGTTGAACTGGTGCTAATACAGGTGTAAT	3057
Qy	246	AAATGACGCTGCACAGCTGCAGATCAATCAAATTTCTTAGAGGGGATTTTTCTTCCATCATTTAG	305
Db	3058	TGCTACAACHTTTCACAGCTGCTGATCATTATCTTCTTTACTCGTGTACTTAAACGTTTAC	3127
Qy	306	CAATATCGATGTACACCAACGGCTTCTGGAGCTGCT	339
Db	3118	TTATGCAAGATGCTAAANAATGCTCAGAGTGTTGCT	3151

RESULT 9  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367

APPLICANT: DORNER, F.  
 APPLICANT: SCHEIFFLINER, F.  
 APPLICANT: FALKNER, F. G.  
 TITLE OF INVENTION: RECOMBINANT FOMLPOX VIRUS  
 NUMBER OF SEQUENCES: 52  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 1800 Diagonal Road, Suite 500  
 CITY: Alexandria  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22313-0299  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/232,463  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/935,313  
 FILING DATE:  
 APPLICATION NUMBER: EP 91 114 300.6  
 FILING DATE: 26-AUG-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 30472/114 INNO  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703)856-9300  
 TELEFAX: (703)683-4109  
 TELEX: 899149  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7218 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 CLONE: pTZ9pl-F15  
 JS-08-232-463-14

Query Match	1.3%	Score 34;	DB 1;	Length 7218;
Best Local Similarity	3.9%	Pred. No. 5.7;		
Matches 13;	Conservative 178;	Mismatches 143;	Indels 0;	Gaps 0;

[illegible][illegible]

```

1342 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1375
Db

```

RESULT 10  
US-08-845-258-10  
: Sequence 10, Application US/08845258  
: Patent No. 6183976  
: GENERAL INFORMATION:  
: APPLICANT: Reed, Steven G.  
: APPLICANT: Lodes, Michael J.  
: APPLICANT: Houghton, Raymond  
: APPLICANT: Sleath, Paul R.  
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
: TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
: NUMBER OF SEQUENCES: 53  
: CORRESPONDENCE ADDRESSES:  
: ADDRESSEE: SEED AND BERRY  
: STREET: 6300 Columbia Center, 701 Fifth Avenue  
: CITY: Seattle  
: STATE: Washington  
: COUNTRY: USA  
: ZIP: 98104  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/845.258  
: FILING DATE: 24-APR-1997  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Makl, David J.  
: REGISTRATION NUMBER: 31,392  
: REFERENCE/DOCKET NUMBER: 210121.426C1  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (206) 622-4900  
: TELEFAX: (206)682-6031  
: INFORMATION FOR SEQ ID NO: 10:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 3701 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
US-08-845-258-10

Query Match	1.3%	Score 32.8	DB 4	Length 3701
Best Local Similarity	47.5%	Pred. No. 9.1		
Matches 97	Conservative 0	Mismatches 107	Indels 0	Gaps 0

Oy	1030	AAATACCTTAAAATGTTGGTGTTCCTTTAAAAATGGAGCCTGGTTTTAAAGCGAAGTGCCCT	1089
Dd	2422	AAATATCTAAGAATGGTATTATTTGTTATGAATATGTAAGTCTCGTAAATGCAGTCTTAT	2481
Oy	1090	TTAAGTCGCAAGCTTCTCTCAGCATGCCAAACTCTAACCTATTATATGCAATTTAGGACG	1149
Dd	2482	ACTCGTAAAAATGATAATGAGCATTAAAGAAATTAGCAGAGATTCATCTGTAATGAAGAAA	2541



	Matches	104;	Conservative	0;	Mismatches	110;	Indels	0;	Gaps	0;
QY	36	AGTACTACGCTTTTATGTCGATTAATTAATAGACTCTGCTACTACCGAAGACTATCGCG	95							
Db	15834	AATATATCTCTGTAATTAATTTGTCAACTCTAGAGAACCCTCATAGACAGCATTAAGATT	15893							
QY	96	TAGCAATAGCTTCGATGGAATACATCAACAAGCTTTCTAGTAAACAATCATATCGCG	155							
Db	15894	AATCAAGCCCTACTAAAGAGAAATTTGAAAAATACAGTTTCTTAATTATACAGTCTCTTC	15953							
QY	156	TACAGATGCGCAACCAATTATGTTTTTAAAGATTCTGATGTATAGAAATGTACCCAAAC	215							
Db	15954	TCCATGGTCCAAAGTTAATGCAAGTTCTGTAATCTGACAGTTAGTATACAAAAAGAAAAA	16013							
QY	216	AGGGAAATCAGTCTACTAGTGTGTTTTAAAT	249							
Db	16014	AAAAAACCAAGCATATGATTTGCTTATATT	16047							

RESULT 4  
 US-08-630-916A-45  
 Sequence 45, Application US/08630916A  
 Patent No. 6011137  
 GENERAL INFORMATION:  
 APPLICANT: Pirozzi, Gregorio  
 APPLICANT: Kay, Brian K.  
 APPLICANT: Fowles, Dana M.  
 TITLE OF INVENTION: IDENTIFICATION AND ISOLATION OF NOVEL  
 TYPE OF INVENTION: POLYPEPTIDES HAVING MW DOMAINS AND METHODS OF USING SAME  
 NUMBER OF SEQUENCES: 124  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: United States  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/630,916A  
 FILING DATE: 03-APR-1996  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: MISROCK, S. LESLIE  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 1101-203  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 896-8864/9741  
 INFORMATION FOR SEQ ID NO: 45:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2052 bases  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA  
 US-08-630-916A-45

```

Db      272 GATCCTCCAGTTCAAGAATAACTGCTACCTCCACAGAAAACAATGAATCTATTCTTCACC 331
Oy      244 AAAAATGACGCTGCAGCTGGAGATCTTAATTCTTAG 280
        | || | | | | | | | | | | | | | |
Db      332 AGTGCAGAAATTGGAATCTGAAGCTGAGAGTATTATAG 368

RESULT          5
US-08-682-517-13
; Sequence 13, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Expression of surface layer proteins
NUMBER OF SEQUENCES: 25
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/682.517
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 3666 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-682-517-13

```

	Query Match	Similarity	1.3%;	Score	34;	DB	2;	Length	3866;	
	Best Local Similarity		45.3%;	Pred.	No.	3.9;				
	Matches	124;	Conservative	0;	Mismatches	150;	Indels	0;	Gaps	0;
OY	66	AGGAGCTCCCTACTACCGAAGAGCTATCGGCTAGCAATAGTTCGTGAGGAACATCATCAAC	125							
Db	2694	AGCAGCTGGTGTTAGTTAGTTGATTAAACAATGCACACTTAGAATTCATTAGC	2753							
OY	126	AACAAGCTTTTCTAGTAAATAACATCATCGGCTACAGATGGCACCAATTATGTTTTAAAGA	185							
Db	2754	TGATGCAGATCTTAATGTAAGTAAGTCACAACTGTTGATCACTGTTTCATTAAAAAGA	2813							
OY	186	TTCTGTAGTTATAGAAAATGTACCACAAACAGGGGAAACTAGTCTACTAGTTGTTTAA	245							
Db	2814	TAGTGCAAAATATTCAATTATCTCTTACATTAGTTGAAACCTGGCTCTAATACAGGTGTAAT	2873							
OY	246	AAATGAGCCTCAGCTGAGATCTAAATTTCTTAGAGAGGGGATTTTCTTCCACATTAG	305							
Db	2874	TGCTAACACTGTTCAGGTGGTAACTTATCTTCTTAACTGCTGGTACATTACAGATTAC	2933							
OY	306	CAATTGCTATGCAACACACAGGCTTCMGAGAGCT	339							
Db	2934	TTATGCAATGCTTAAAAATGCTGCAGAGGTGTTGCT	2967							

[illegible]





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 03:44:51 ; Search time 268.88 Seconds  
(without alignments)  
1778.488 Million cell updates/sec

Title: US-09-446-677B-7

Perfect score: 2526  
Sequence: 1 ATGAGAGTTCACCTCCGCTT.....GTAGCAATCAATTG 2526

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCrus.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	38	1.5	7218	1 US-08-232-463-14	Sequence 14, Appl
2	38	1.5	20303	1 US-08-370-975B-6	Sequence 6, Appl
3	38	1.5	26764	1 US-08-370-975B-1	Sequence 1, Appl
4	35.4	1.4	2052	3 US-08-630-916A-45	Sequence 45, Appl
5	34	1.3	3666	2 US-08-682-517-13	Sequence 13, Appl
6	34	1.3	3666	2 US-08-682-517-14	Sequence 14, Appl
7	34	1.3	4197	2 US-08-682-517-7	Sequence 7, Appl
8	34	1.3	4197	2 US-08-682-517-8	Sequence 8, Appl
9	34	1.3	7218	1 US-08-232-463-14	Sequence 14, Appl
10	32.8	1.3	3701	4 US-08-845-258-10	Sequence 10, Appl
11	32.8	1.3	3701	4 US-08-990-571-10	Sequence 10, Appl
12	32.6	1.3	5470	1 US-08-441-139-12	Sequence 12, Appl
13	32.6	1.3	5470	6 5196523-5	Patent No. 5196523
C 14	32.4	1.3	1102	4 US-08-821-994-86	Sequence 86, Appl
15	32.4	1.3	1200	1 US-08-357-264-2	Sequence 2, Appl
16	32.4	1.3	1200	1 US-08-672-514-2	Sequence 2, Appl
C 17	32.4	1.3	1434	4 US-08-821-994-62	Sequence 62, Appl
C 18	32.2	1.3	1712	4 US-08-809-326A-8	Sequence 8, Appl
C 19	32.2	1.3	1954	4 US-08-809-326A-13	Sequence 13, Appl
C 20	32	1.3	4507	2 US-08-568-459A-3	Sequence 3, Appl
C 21	32	1.3	4507	2 US-08-487-826B-3	Sequence 3, Appl
22	31.8	1.3	834	3 US-08-633-993A-12	Sequence 12, Appl
23	31.8	1.3	834	3 US-08-844-188-12	Sequence 12, Appl
24	31.8	1.3	1152	3 US-08-844-188-42	Sequence 42, Appl
25	31.8	1.3	2132	3 US-08-844-188-39	Sequence 39, Appl
26	31.8	1.3	4084	2 US-08-568-459A-1	Sequence 1, Appl
27	31.8	1.3	4084	2 US-08-487-826B-1	Sequence 1, Appl

28	31.8	1.3	4084	6 5198347-5	Patent No. 5198347
29	31.6	1.3	674	4 US-09-040-984-40	Sequence 40, Appl
30	31.6	1.3	4565	3 US-08-776-265-2	Sequence 2, Appl
31	31.6	1.3	5084	1 US-08-306-691B-21	Sequence 21, Appl
32	31.6	1.3	5084	5 PCT-US93-06251-25	Sequence 25, Appl
33	31.6	1.3	5319	1 US-08-169-927-1	Sequence 1, Appl
34	31.4	1.2	372	3 US-08-844-188-40	Sequence 40, Appl
35	31.4	1.2	1902	3 US-09-041-991A-7	Sequence 7, Appl
36	31.4	1.2	1902	3 US-09-041-991A-9	Sequence 9, Appl
C 37	31.4	1.2	3164	1 US-07-723-002C-3	Sequence 3, Appl
38	31.4	1.2	3989	1 US-08-327-494A-1	Sequence 1, Appl
39	31.4	1.2	3989	1 US-08-327-494A-3	Sequence 1, Appl
40	31.4	1.2	3989	5 PCT-US95-13659-1	Sequence 1, Appl
41	31.4	1.2	3989	5 PCT-US95-13659-3	Sequence 3, Appl
C 42	31.4	1.2	9793	1 US-08-470-202-56	Sequence 56, Appl
C 43	31.4	1.2	9793	1 US-08-471-770-56	Sequence 56, Appl
C 44	31.4	1.2	9793	2 US-08-468-059-56	Sequence 56, Appl
45	31.4	1.2	176373	3 US-09-128-155-17	Sequence 17, Appl

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEFFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
CLASSIFICATION: 435  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-9300  
TELEFAX: (703) 683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-F1s  
US-08-232-463-14

---

**THIS PAGE BLANK (USPTO)**



```

Db 1918 tccctaaagccitgtgggtctcttcttctgcatgtccgctccatccagagccatagaccg 1977
QY 1599 AGGTACTGAGAGGTCCTCTTACGAAAGATTTTGGTTCAGGACTTTCGAATGTTT 1658
Db 1978 gagcacaattcgtctctctcgtcaacaatttgygtatccaggaatcgcgactttt 2037
QY 1659 GCATAGAGACGGTCGTGAAATTCAGAAATTCCTCATGTGATGAGAGGTCGTCTACT 1718
Db 2038 gcatgagaatcagaagaagaaacacagtagtatcgtaactctcagcgsggtatgcat 2097
QY 1719 AGGTGCTACACAGAGATCCCGGTCGTGATACCTTGTCTGCGTTTTCGTCAGCTCT 1778
Db 2098 aggaagagagatctctcagcgctctcgaattcttctaattcttgccttcttgcagctt 2157
QY 1779 TCGCGCTGACAAAGACTTACTTATGATACCAATTCGAAAGACTAGCGAGATCTT 1838
Db 2158 tggctacgacaagagaccatcttgygtacagaacacacatctgtatatgacagggcaat 2217
QY 1839 ACGTTTGACAGCAGATGCTTCCCTATACTCTGTGAGTATCCCTTWTAGAGAGGAGG 1898
Db 2218 ggttacgcgacac-----ctcgagagajctaa 2244
QY 1899 ACTCCGCGAGATCCTCTTCTTATCTTCCAAAGACTTCGCCGTCTTCTTATGGCA 1958
Db 2245 gacctcgtacagatcttctcaggaattctgactccactcttctcctcaatgctcg 2304
QY 1959 GCTTACTGCGGCGCATGCGATCATGCAAGACGTCCTACCCGCCGCCGCCGCCGCC 2018
Db 2305 gtttctcttgcacatacgcgaataacatgacacacaagt-----acgc 2349
QY 2019 GACGCTCTGACGATCATCTTCTTGGGAGATATGCTGGGCTGAGAGCTGGGAAC 2078
Db 2350 tggctatctctcgttaaggaagctggggaatgtagctcttgatagatgtgagag 2409
QY 2079 TCGAGTGTGCTGTGAAAATACCAACGCGCAGAGATTTTCCAGAGTAACTCCATTTGT 2138
Db 2410 agctatcccgtagtctgctcaggaagctgcttgygtgagatcccaacagcatttct 2469
QY 2139 AAAAGTCAAGCTGTTTACTCGCGCCAGATAGCTTGTGAACACTAGAGCTATCAGTGC 2198
Db 2470 aaacctagaagatgatatgacacatcagaatgacttaaggaagaaacgcagcagaagccg 2529
QY 2199 TGAATTTAGTATTCGATTCGATCTTATTAACCTTGGATTCCTTGAATCAAGTTAGAGA 2258
Db 2530 tctcttccaaaggaagacctctcactcaactagcgtctcctgtaggataaaattgagaa 2589
QY 2259 ACGGTTTGCAGAGCAATATTAATCATGTGTAGAGATGATTCGCCAGATGTTTGTGTAG 2318
Db 2590 attctccgaataagctacgtatgactctcactagcttactcgtccgactgtattgtaa 2649
QY 2319 TAAACCCAAATGATGACGACTACCTTCTTCAACCAAGAGAGTTGGAGAACCAAGGTTTC 2378
Db 2650 tgcatacagctgcacagaacactctatgcttctcgtggatcttctgtagacatgtygtac 2709
QY 2379 GAACTTAGCAAGACAGCTGATTTGTTTCAGGCTTCAGGCTTTCGATCTTGGAGACTGC 2438
Db 2710 aagcttgcataagacaagcctctcttctgtagcgtcgtgaaatcatcatgcttctgctcaaa 2769
QY 2439 AGGAGAGCTTTTGGGAATTTGGCTTGAATGGCGGGGATCTTCCGATGATTAATGT 2498
Db 2770 ctttgaagttctcagctcagctcgaatcgaatgctgcgaaggtcttctcgtagactat 2829
QY 2499 AGATGCGGTAGCAAAATCAAAATTTTA 2525
Db 2830 cgatctctgagagaagatctcgatcttta 2856

```

```

DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp9 DNA.
XX
KM Omp9; outer membrane protein 9; surface exposed protein; antigen;
KW infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX
XX Chlamydia pneumoniae.
XX
PN W09858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98W0-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.
XX
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Myglind P;
XX
DR WPI: 1999-105610/09.
DR P-PSDB: AAM06822.
XX
PT Species-specific test for identifying mammals infected with
PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT these proteins
XX
PS Claim 6; Page 55-56; 115pp; English.
XX
CC This DNA sequence codes for the novel 96.7 kDa surface exposed
CC protein Omp9 (see AAM8422) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see AAM06816-27) encoding
CC Omp9-Omp15 proteins (see AAM8417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
CC the other, and encode polypeptides of about 89.6-100.3 kDa and
CC about 56.1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.
XX
SQ Sequence 2757 BP; 770 A; 564 C; 642 G; 781 T; 0 other;

Query Match 11.7%; Score 295.4; DB 20; Length 2757;
Best Local Similarity 49.6%; Pred. No. 7,2e-76;
Matches 995; Conservative 0; Mismatches 931; Indels 81; Gaps 6;

QY 543 CTTGAGATGCGCAACATATAGTCCCTTCTTATTTGGAATATGTTCTTCAACAGCTGG 602
Db 807 ctgactcttccaggaagaagagagcctgcatcttctcgataaacacagcaggaatagtg 866
QY 603 CGAGCGATTCAATACCAAAACCTACACTATCTTCTGCT---GGGAAACCTATTTGA 659
Db 867 agtgcgattatataccaataaattggtgtatccctcagaacgagaggggtgtattcttc 926
QY 660 GGGGATATAGCGCGCTGCTGTGTAAGAGAGTGTATTCGGATTCGAGACTCTGG 719
Db 927 taacacaagctgcgaatgctactcctaagagaggggcaatgctgactctagatctcgg 986

```



CC in the diagnosis of *C. pneumoniae* infection in mammals. The  
CC nucleic acids and proteins can also be used in the immunization of  
CC mammals, the nucleic acids being particularly useful as DNA  
CC vaccines for effecting *in vivo* expression of antigens. The  
CC vaccines may also prevent atherosclerosis and bronchial asthma,  
CC which are possibly associated with *C. pneumoniae*.

XX Sequence 2815 BP; 790 A; 575 C; 626 G; 824 T; 0 other;

Query Match 12.2%; Score 309.4; DB 20; Length 2815;  
Best Local Similarity - 50.0%; Pred. No. 5.9e-80;  
Matches 1028; Conservative 0; Mismatches 951; Indels 78; Gaps 7;

```
QY 509 CAGGATGGGCGAGCAATTAATTGACGGCCCTTGGAATCGCAACAACTAATGCTCC 568
DB 770 caggaatggaagcctcttcgagatgcagatgtaccacatctggaaacagaatg 829
QY 569 TTTCTTTTATGGAATAGCTTTCACACGTGGCGAGCGATTACCAAAAACCTCA 628
DB 830 taacttcaggaacacagctgtagctaatgycggagccattatgtctaagaactta 889
QY 629 CACTACTTCTGTGGGAAA-----CTCTATTTCAGGGGAATACAGCCCTACGG 679
DB 890 cactggcttcgagggggggggtatctctcttcttaacaatagctcaaggtacaa 949
QY 680 CTGCTGTAAAGGAGGTGATCGGATTGACAGACTGTGGACACCATCATCTTCGGAG 739
DB 950 ctgagcgaatagtgagcaatctcatctcagctgagagatgtagcttcttcagag 1009
QY 740 ACAGTGGACATTAATCTTTGGAAGCAATACGATAGAGCTA-----CAGAACCGTCT 793
DB 1010 aagcagggagcattactcctaagtgatgcatgttgtaactacacacaactacaa 1069
QY 794 CTCTATGCTCTATTGATTAGAACACTAGCGCTAAGTAACTGCGTTACGTCCCAAG 853
DB 1070 aagaatcttattgacatagatctcactgcgaagaatcaagaatlaagtcataatcgt 1129
QY 854 GACATACGATATATCTTTTATGATCGATTACGTATACAGAGATGACATGCTGCTGATG 913
DB 1130 ggcatagcatacttcttctacagatccgatactgcttaatacgcgtgagatcttcaagata 1189
QY 914 CTCTCAATTAATTAATCCCTGATATGAGATTAACAAGATTAACGGGAACCATAGTCT 973
DB 1190 cttaaatccataaagcgtcgtacgtatagatagattagtggtgcagtctt 1249
QY 974 TTTCTGAGAGAACTCAGCGAGGAGCAAGCTTAAGATAGAGAAGAACCGACTCTTAAT 1033
DB 1250 ttctctggtgaagcctctcgaagatgaagcaaaagttcagagacaacttactacgc 1309
QY 1034 TACTTCAAAATGTTGCTTTTAAATAATGACCTGATTTTAAAGATGATGCTTTTAA 1093
DB 1310 tgaagaagcctgttaactcttaactgcaggaatattagtaactaaagctgtgtactcctcg 1369
QY 1094 GTGCGAAGCGTTTCTCTCAGAGATCAAACTTAAGTTATGATTAAGGAGCTCGT 1153
DB 1370 atcagaaagccttactacgacgcgggttccctctgtatlatiagatgcyggaacaagt 1429
QY 1154 TGGTTGCAAAACACGAAAGATGAGATTAAGCAATTTGAAATTAATATAGATCTCTCA 1213
DB 1430 taagaacagtaagaagaggtcactttaaagagttccacttctcgttagactctttag 1489
QY 1214 GGAACGGAAGAAAGATTAACACTAGTGTGCCACAGCTCAGAAAGATTAATTCGATGATC 1273
DB 1490 gggagggtaagaagatgttaattgtcgtcttcgcagcaagtaaaatgttagcccttagtg 1549
QY 1274 GTCTCTGTACTGGCAATAGGATGAGATTTTATCAAAAATGGCTTTTGAATGAGG 1333
DB 1550 gtccgactctctctttagtaacaaaggaatgtcttaaaatcagactttagaanaa 1609
QY 1334 ACGATTCGATGATGGGATTTCTGAGTTAGATCTGGAAGAAAGCATCGATTTCTCGAG 1393
DB 1610 ctcaagaacttcaattgtgagagctctcgtcgttactgcaacaactaagatgcttc 1669
```

```
QY 1394 ATTCTCGAGTATTAATGCTGTACATTCGCCGATGCTATCGAGGAAGTGACATCA 1453
DB 1670 cagcggtctccacagtaagaactctctacgatacgtggtatacaagttacttgggaaatga 1729
QY 1454 ATTTGCTACTGAT-----GATTAAGAAGCTACGCTTTCTTGGGCA 1495
DB 1730 ctgggttgatgataccgcaagcaactccaagaactaaagataagcagcgacattagcttgaca 1789
QY 1496 AGCAAAAGTTTATCCACTGCTGACGACGAGGCTCCGTTAGTCTTATCTTTGGG 1555
DB 1790 atacaagctaccctccgaatcccgatcgatcaagggaccttagttcctaagccttgg 1849
QY 1556 GTTCTTTTATAGATGTTCTCCCTTCCAAAATTTTATAGCTAGTAGTGAAGGTCGTC 1615
DB 1850 gatcttcttcagacatccaaagatccaaggtatcaaggtatagaagaaatgtttagacctt 1909
QY 1616 CTTACGAAAAGAGATTTTGGGTTGACAGCAATTCATATGTTTGCATAGAGCGTCTG 1675
DB 1910 gtccagatcgagctcttggtcgtcgaggatcgcaattctttagataagaataagaag 1969
QY 1676 AAATCAAGAAATTCGTCATGTAGTGGAGGTGCTGTAGTAGTGTGATGACGACGAG 1735
DB 1970 gggaaaaacgcgaatacgcgtcaataatctggtgatalatgatacggaaggtgcagcaaa 2029
QY 1736 TGCCGGGTGATACCTTGCTCTGGGTTTGTGCTCAGCTCTTTGGCGGTGACAAAGACT 1795
DB 2030 ctgtgtctgaataactaatatagcttgccttcttgccaactcttgttagcgataaagatt 2089
QY 1796 ACTTTATGAATACCAATTTTCGAAAGACCTACGACAGATCTTTACGTTTGCAGACAGAT 1855
DB 2090 tcttagtcgataaatactatcactatgatacgaagagccttcttataccacaacatla 2149
QY 1856 CTTCCTTACTCTGTGGGAGATCTCTTTTATGAGAGGAGAGACTCCCGAGATCTCTGT 1915
DB 2150 cagaatgtatggtgtatcagaatgtgtctcttagataa-----ac 2188
QY 1916 TGCTTATGTTTCCAGAGCTTGCCGCTGCTTTATGATGGAGCTTATGACGCCATA 1975
DB 2189 ttccgtgccttgtagatcaataaccctcgtttlaagaaggagcgtcgttataagcaag 2248
QY 1976 CGGATCATGATGAAGACCGAGTCTTACCCCCCCCCCGGACGCTCTGACGGATC 2035
DB 2249 tcaagtaatgatctgaaagaagaatla-----actgcgtatctcgtatgta 2293
QY 2036 ATACTTCTGGGAGGATATGCTGTGGCTGAGAGCTGGGAATCGATGCTGTGAA 2095
DB 2294 aaggttcttgagggaataatgtctttaaacaatggttggagct---tcttcatactt 2350
QY 2096 ATACGACGCGAGAGATTTTCCGAGAGTACACTCCATTTGTAAGAATCCAAAGCTGTT 2155
DB 2351 atccgtataccgcatgtgtttagatactatgctcactatcactcaacatcgaatctgacct 2410
QY 2156 ACTCGCGCAAGATAGCTTTGTTGAAGTACGACTATCAAGTGTGATTTTATGATTCG 2215
DB 2411 atatacgtacagacactctcgcgagaagaaggtacagaagaagaatcttctttagacaga 2470
QY 2216 ATCTTATTAACCTTGCGATTCCTCTGTAATCAAGTTAGGAAGCGTTG-----CAG 2269
DB 2471 acccttcaattatctcttgccataggggtgaggtctttagaaggtctctcgtatgtaagt 2530
QY 2270 AGCAATATTAATCATGTTGTAGGAGATTAATTCACAGATGTTGTGCTAGTAACCCCAAT 2329
DB 2531 accttcttaatgatactgacttatactatgcttactgctcacttactccgaatgctccaat 2590
QY 2330 GTACGACTACCTTACTTCTTCCAAAGGAGATTTGGAAGACCAAAAGTTTGAAGTACGA 2389
DB 2591 gacactaagcacttgaactcaacagcgagcctcttggaanaactatgcaaaataacttgcag 2650
QY 2390 GACAGGCTGATGATGTTTACGGGCTGAGGCTGTTTGCATCTTTGGAGGCTGACGACGACT 2449
DB 2651 gaaagccttgcaagtgtgtgacagcagatcactagccttctcctcetaagttagaagtc 2710
```



```

Db 99743 tccgattctcttggtaaccaagggaatgcttatgaataatcagccttagagaanaac 99802
Oy 1335 CCATTCCCATGATGGGATTTCTTGAGTTGATGCTGGGAAAGACATCGATTTCTGCGA 1394
Db 99803 tcaagacttctcaattgtgcagctctcgtcctggttactcacaacataagatgltcc 99862
Oy 1395 TTCGCGCATTAATGATGCTAGACATCTCCGATGCTATGAGGAAAGTGCATCAATCA 1454
Db 99863 agcggttctcagtagaactcctcagcactatggtatcaagttacttgggaatgac 99922
Oy 1455 TTGGTCTACTGAT-----GATAGAAGACTACGCTTTCTTGGGCAA 1496
Db 99923 ttgggtgatgataccgcaagcactccaagaactaagcagcagcattagcttgaccaa 99982
Oy 1497 GCAAGATTATCCCATGCTGAGCAGAGAGCTCCGTTACTTCTATCTTTCTTGGGG 1556
Db 99983 tacaggtacacttcgcaatcctgagcgtaaggaaccttagcttctaataagccttgg99 100042
Oy 1557 TTCTTTATAGATGTTGCTCCCTTCCAAATTTTATAGAGTAGTACTGAAAGTCTCC 1616
Db 100043 atcttttccagacatccaagcagatccaagtgctcatagaagaagtgcttggactcttg 100102
Oy 1617 TTACGAAAAGATTTTGGGTTGCGAGCATTTCCAAAGTTTTCATAGAGACCGCTCGTA 1676
Db 100103 ttccagatcgaagctctggtcgtcg99agtcgccaattcttagaataagaataagaag 100162
Oy 1677 AAATCAAGGAAATTCGTCATGAGTGAGTGAGTGTCTAGTAGTGTCTACACAGAGAT 1736
Db 100163 ggaataaacgcaataacgctcaataaactcgtcgtatgatactcgaagtgtagcgaac 100222
Oy 1737 GCCGGTGGTATACCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1796
Db 100223 ttgttcgaaactaatatagcttgccttgcactcttgccaactcttggtagcctaagaatt 100282
Oy 1797 CTTTATGAAATACCAATTTGCGAAAGACCTACGAGATCTTTAGCTTTGACGACGATGC 1856
Db 100283 ctagtcgtaaaaatacactaactatgactcagcaggaagccttctatataccaacattac 100342
Oy 1857 TTCCCTATACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1916
Db 100343 agaattagtagtggtcatalagtgltcctatgataa-----act 100381
Oy 1917 GCTTATGTTTCCAGACTGCTGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1976
Db 100382 tccgtgctcttgaggtcatalaaccctcgtttagaagggcagctcgcttagcagaagt 100441
Oy 1977 GGATCATCGCATGAGACGAGCTCTACCCGCCGCCGCCGACGCTCTGACGAGATCA 2036
Db 100442 cagtaatgactcgaagaagaat-----actgctatcctgaggtgaa 100486
Oy 2037 TACTTCTTGGGAGGATATGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2096
Db 100487 aggttcttg99gaataatgcttctaacaatglttgg99aact---tcttctacttcta 100543
Oy 2097 TACCAGCGGAGAGATTTTCCAGAGTACATCCATTTTAAAGTCAAGCTGTTTA 2156
Db 100544 tccgtaataactcgtatgltttagtaactatgctccataactcaactgaactcgtacta 100603
Oy 2157 CTCGCGCCAGATAGCTTTGTTGTAAGTACGATACGTCGTATTTAGTATTCGA 2216
Db 100604 tatacgcagaagcgtctcgtcgagaaggtacagaagaagcattcttggtagcagaa 100663
Oy 2217 TCTTTATTAACCTTGGCATCTCTTGGAAATCAAGTATAGAAAGAGGTTG-----CAGA 2270
Db 100664 ccttctaataactcgtcgtcgtatagaaggttgaagttgagttcgtatgtaagaa 100723
Oy 2271 GCAATATTAATCATGTTGAGAGATGTTTCCAGATGTTTGTGATGTTTATCCCAATG 2330
Db 100724 ctttcttatagacttgacttataccctatgctccgacttataccgcaatgctcccaatg 100783
Oy 2331 TACGACTACCTTACTTTCCAAACCAAGGAGTTGGAAGACCAAGGTTGCAACTTACGAG 2390

```

```

Db 100784 cactacagcacttgtaalcagcgagcctcttgg99aactatagccaataacttagcacg 100843
Oy 2391 ACAGGCTGTAATGTTGTCAGGCTCAGCTTTGCAATCTTGGAGCTGCAGAGACTTTT 2450
Db 100844 acaagccttcaagaagtgctgcagcagcactacagcctctctccatagtttgaaggtc 100903
Oy 2451 CGGGAATTTGCTGCTGAATGCGGGGATCTTCTGTAATATATATGATGCGGGTAG 2510
Db 100904 cggcaggttcttgaagttcgtggtcctcctcaaggaattataatgtagatcttgg99 100963
Oy 2511 CAAATCAAAATTTAG 2526
Db 100964 taagttccaattcag 100979

RESULT 13
AAK06817
ID AAK06817 standard; DNA, 2815 BP.
XX
AC AAK06817;
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp5 DNA.
XX
KW Omp5; outer membrane protein 5; surface exposed protein; antigen;
XX infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX Chlamydia pneumoniae.
OS
FH Key Location/Qualifiers
FT CDS 1..2787
FT /tag= a
XX
PN M09858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98WO-DK0266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
XX (CHRI/) CHRISTIANSEN G.
XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX Myding P;
XX WPI: 1999-105610/09.
XX P-PSDB: AAM88418.
XX
PT Species-specific test for identifying mammals infected with
PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT these proteins
XX
PS Claim 6; Page 42-43; 115pp; English.
XX
CC This DNA sequence codes for the novel 97.2 kDa surface exposed
CC protein Omp5 (see AAM88418) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see AAK06816-27) encoding
CC Omp4-Omp5 proteins (see AAM88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
CC the other, and encode polypeptides of about 89.6-100.3 kDa and
CC about 56.1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp5 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used

```

```

Db 2341 cctttgtcttcttccacaccacacccctcttcaagacccttcaatccacagatgaagtc 2400
Oy 2146 CAAGCTGTATTACCGCGGCAAGATAGCTTGTGAACATAGAGCTATCAGTGTATTT 2205
Db 2401 gaaatggttataatcacaataaagctcttcgaagctagtagtgccgtggtt 2460
Oy 2206 AGTATGTCGCACTTTATTAACCTTGGGATTCCTTGAATCAAGTATGAGAAGCGTTT 2265
Db 2461 agtatgtgaagctgtcttaacctctgactctctgtgtgtgtgcgaatctgtcgaagggt 2520
Oy 2266 GCAGAGCAATAT-----TATCATGTTGAGCGAGTATCTCCAGATGTTGCTGACT 2319
Db 2521 atcgagagatctctacacatgatctctcagagatcttcttccgatgtcatalcgaac 2580
Oy 2320 AACCCCAATGATGAGCTACCTACTTTCACCAACCAAGSAGATGAGACCAAGTTGCG 2379
Db 2581 aaatcccaatctacagagactcttgatgagccagactcttggaatctgcgtgtgc 2640
Oy 2380 AACTTATGACAGACAGGCTGTATTTGTTTCAGGCTCAGTTCGATCTTGGAGCTGCA 2439
Db 2641 aatcttcaagaacagcatttctactgagggtagaacaactaagctcacaactcaat 2700
Oy 2440 GCAGAGCTTTTGGGAACTTGTGCTTGAATGGCGGGAATCTTCTGATGATTAATGTA 2499
Db 2701 tctgagctcttcgacatcagctatgagactcgtgtgacttcaagaaactacaatgta 2760
Oy 2500 GATGGGATGACCAATCAATTT 2522
Db 2761 gatgtgtacacaaactcgaatt 2783

```

RESULT 12  
AAC81914  
ID AAC81914 standard: DNA; 273254 BP.

AC AC81914:

DT 27-FEB-2001 (first entry)

DE Chlamydia pneumoniae genome DNA.

KX Genome; diagnosis: vaccine; ds.

OS Chlamydia pneumoniae.

PN W0200027994-A2.

PD 18-MAY-2000.

PF 12-NOV-1999; 99WO-US26923.

PR 12-NOV-1998; 98US-0108279.

PR 08-APR-1999; 99US-0128606.

PA (REGC) UNIV CALIFORNIA.

PI Stephens R, Mitchell W, Kalman S, Davis R;

DR WPI: 2000-376516/32.

PT Isolated nucleic acid for use in diagnostic and analytical methods

PS encodes genomic sequence of Chlamydia pneumoniae -

XX Claim 2: Page 128-320; 320pp; English.

CC This invention describes a novel nucleic acid (NI) encoding a Chlamydia pneumoniae protein (PI), given in the specification. The isolated nucleic acid is useful for diagnostic and analytical methods, such as, hybridization-based assays or amplification-based assays. The protein may be used for diagnostic purposes, for their enzymatic or structural activity, or as a vaccine. The invention also describes (1) a probe comprising a hybridizing fragment of NI; (2) an isolated nucleic acid

CC (N2) that hybridizes under stringent conditions to NI; (3) an expression cassette comprising NI under the transcriptional regulation of a transcriptional initiation region functional in an expression host, and a transcriptional termination region; (4) a cell comprising an expression cassette of (3) as part of an extrachromosomal element or integrated into the genome of a host cell as a result of induction of the expression cassette into the host cell, and the cellular progeny of the host cell; (5) a method for producing a PI comprising growing a cell of (4) where the protein is expressed and isolating the protein free of other proteins; (6) a purified polypeptide composition comprising at least 50 weight % of PI; and (7) a monoclonal antibody binding specifically to the peptide of (6).

CC Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

Query Match 12.4%; Score 313.6; DB 21; Length 273254;  
Best Local Similarity 50.1%; Pred. No. 4e-80;  
Matches 1030; Conservative 0; Mismatches 949; Indels 77; Gaps 7;

```

Oy 509 CAGGAGATGCGGAGCAATTAATGTCAGAGCTCCTTGAAGATCCGAACAAATTAAGTCCC 568
Db 98963 caggaaatggaagagctcttctgagatgcccagatgtaccatctggaatcagagt 99022
Oy 569 TTTCTTTTATGGAATTAATGCTTCAACAGTGGCGAGCATTCATACCAAAACCTCA 628
Db 99023 taacttccagaaaccaaagctgtatgactaagcgagccattatgttaagaagctta 99082
Oy 629 CACTATCTTCTGTGGGGAACCTATTTCAGGGGAATACAGCCG-----TAGGCG 680
Db 99083 caatgctctcgggggggggggggagatctcttcttcaacatatagtccaaagatccac 99142
Oy 681 TGCTGTAAGAGGTCGATGCGATTCGAGATTCGACACTCTCCATTCATTTCTGAGA 740
Db 99143 tgcaggtaatgtgtgagccattctctactgacgtgtgaggtgtgagcttcttcagca 99202
Oy 741 CAGTGGCAGATTATCTTTGAAGGCAATAGATAGAGCTA-----CAGAAACCGCTTC 794
Db 99203 agcagggagacattacccttcaatggaatgcatgttgtaacacacaaactacaa 99262
Oy 795 TCATAGTCTATTGATTTGGAACATGAGCTTAAGTAATGAGTTCGTCGCGCAAG 854
Db 99263 aagaaatctatgacatgactgactcgcgaagacacgaatcttgcgtacatctcgt 99322
Oy 855 ACATACGATATATCTTTATGATCCGATTAATGATGACAGATGACATCTGTTGCTGATC 914
Db 99323 gcatgcatcttcttctacgactctactctctactaataagcgtgagatctcaagatc 99382
Oy 915 TCTCAATTAATTAAGCCCTGATCTGAGATTAACAAAGATATACGGAAACATAGCTTT 974
Db 99383 tttaaatctcataaagcctgtagcaggtatagtagacagattatagtgagtgatgtt 99442
Oy 975 TTTCTGAGAGAGCTTCACGAGGCAAGCACTTAAGATGAGAGAACCCGACATCTTAAT 1034
Db 99443 ttctgtgaaagctctcgtgaagatgaacaaagcttgagacaaactcactcagct 99502
Oy 1035 ACTTCAAAATGTTGCTTTTAAAAATGAGCTGATGTTTAAAGGATGATGCTTTTAA 1094
Db 99503 gaagcagcctgtaactctactacgcaggaatctagtaactaaagtggtcactctga 99562
Oy 1095 TCGGAACGGTTTCTTCAGAGATCAAACTTAAGTTGATTAATGAGACGTGTT 1154
Db 99563 tacgaagaagcttactacagccggttcctctgttattatgagtgagggaacaagct 99622
Oy 1155 GGTTCGAACACCGGAAGATAGATTAAGATTAATGAAATTAATATAGACTCTCTCAG 1214
Db 99623 aaaaagcaagtaacagagaggtacttaacttaacaggtcttccactcgtgaactttag 99682
Oy 1215 GAACGGAAAAAGATAAACTCAGTGCACAGCTCGAAGATATTCGTATAGATG 1274
Db 99683 cgagggttaagaagtgtaatgtcgtcttcgcagcaagtaaaatgtgccccttagtg 99742
Oy 1275 TCTGTGTACTGGCAATTAAGCATGAGATTTTATCAAAATGCGCTTTTGAATGAGGA 1334

```

PT asthma -  
XX Claim 2: Page 140-141; 174pp: English.  
XX  
CC This sequence encodes a Chlamydia antigen of the invention, designated  
CC CPN100654. The nucleic acids (and their complementary sequences) may be  
CC used as diagnostic agents for detecting the presence of nucleic acids  
CC encoding Chlamydia antigens in samples according to standard methods,  
CC and therefore, for diagnosing Chlamydia infections. For example, they may  
CC be used as primers and probes for diagnostic polymerase chain reaction  
CC (PCR) assays. Antisense sequences may be used to down regulate  
CC expression of the proteins and may be used to treat infections. The  
CC nucleic acids may also be used to produce the protein antigens they  
CC encode according to standard recombinant DNA methodologies. The  
CC proteins may then be used as antigens for the production of antibodies  
CC (i.e. as vaccines) for preventing infection by Chlamydia. The  
CC antibodies may also be used as diagnostic reagents for detecting  
CC infections. Chlamydia is a pathogen implicated in the development of  
CC (for example) community acquired pneumonia, upper respiratory tract  
CC disease (especially bronchitis and sinusitis), asthmatic bronchitis,  
CC adult-onset asthma and acute exacerbations of asthma in adults.  
CC  
SQ Sequence 2784 BP; 803 A; 644 C; 573 G; 764 T; 0 other;  
  
Query Match 12.9%; Score 326.6; DB 21; Length 2784;  
Best Local Similarity 50.0%; Pred. No. 5.5e-85;  
Matches 1092; Conservative 0; Mismatches 1014; Indels 77; Gaps 8;  
  
QY 396 TTTTCTTAATCCCAAGCAAGTACAGTACATGATGATGAGAGCTATCAATGTTAAAG 455  
DB 622 ttatgttagctctctatctacataagctctagtcagagcgctcatgattgtagg 681  
QY 456 GAATTTAAGCTTATGATATGATATGATGATGATGATGATGATGATGATGATGATGAT 515  
DB 682 caagctgatactatcgacaacaattctatcttggaaaggaatgcaagaaataac 741  
QY 516 TGGCGAGCAAT-----AATTGTCAGGCTCTTGAAGTCGCAACAA 560  
DB 742 tggcgagtgagatctgcacaacagcgagtgatccctgaaactgatactcttaaca 801  
QY 561 TAAATCCCTTTCTTTTATGGAATATGATCTTCAACAGCTGGGAGCGATGATACCA 620  
DB 802 taagactctgactcttgcttcaacagtagaagaacagcggtggtccatccatgtaa 861  
QY 621 AAACGTACACTATCTTGTGGGGAACCTGATTTTACAGGGAATACAGCGCTACGCG 680  
DB 862 aaagctagcccttccctctgagagcttcaagagttcttaagaataatgctcatcagc 921  
QY 681 TGCMTGTAAGAGAGGTGCTATCGGATTCGACGCTGCAACCTTATCCATTTCTGAGA 740  
DB 922 aactccaaagggggtgctatcagcatcgatgcctcagaagagctcagcttcttcaga 981  
QY 741 CAGTGGCAGATATCTTTGGAAGCAATAGATAGAGTACAGCAACCTCTCT----- 795  
DB 982 gaaagaaataaactcttgtaagaataacaccttaacaacaacggaaagtaaccgatactc 1041  
QY 796 ----CATAGCTGATGATTTAGAACTAGCGCTAGATTAAGCTGCTGCTGCGCA 851  
DB 1042 taaagctaaagctacacataagaaagtaacggaataatcaagaaattacagggctgtaa 1101  
QY 852 AGGACATGATGATATCTTTATGATCGATTCGATACAGCAACCTCTGCTGCTGTA 911  
DB 1102 aaatcaataaattctctctatgataccatctcagaagaacctcat-----caga 1155  
QY 912 TGGCTCAATATTAATAGCCCTGATGAGATTAACAAGATTAACGGAACCAATAGT 971  
DB 1156 cgtattgaagaataaataagcgctctcgaggagctctcaatcacaataaagaacgattcc 1215  
QY 972 CTTTCTTGGAGAGAGCTACGAGGAGCAAGCTAAAGATGAGAGACCGCACTTTAA 1031  
DB 1216 attctctggaacacccataacagcagatgaacttaagtctgcacaatttaaatcttc 1275

QY 1032 ATTACTTCAAAATGCTGCTTTTAAAAATGGGAGCTAGATTAAAAAGCTGATGCTTTT 1091  
DB 1276 attcagcagccagctccctatcccggaagaagtattcttcaaaaaaggagcacttc 1335  
QY 1092 AAGTGGCAAGGTTTCTCTCAAGATGCAAACTTAATGATTTAGATTTAGGAGCTC 1151  
DB 1336 agagaagcagagctctctcctaagagcggttctctcctcgatcagtgatcagaagac 1395  
QY 1152 GTTGGTTCGAACACGGAAGTATCGAGTTAACGAAATTTGGAATTAATTTAGCTCT 1211  
DB 1396 attcaactcaagctgaggagattacaacaacgaacctgaagtaacatgltgactctc 1455  
QY 1212 CAGGAACGGGAAAAAGATTAACACTAGTGTGCGCACACTCAGAAAGATTTGCTATAGA 1271  
DB 1456 aggtcttaagcagccgctcagcctaacaagaagaagtgcttcaataaagtgatcgtatc 1515  
QY 1272 TCGTCTGTTGACTGGCAATTAAGCATGAGATTTTATCAAAATGGCTTTTGAATGA 1331  
DB 1516 tgggaagctcaacgtatgatatggaaggaaacatttaagaagtacatagttcagcca 1575  
QY 1332 GGACATTTCT-----ATGATGGGATTTCTTGAGTTAGATGCTGG 1570  
DB 1576 tgaccagctctctctctatlaaanaatacaggttgatgctgattgatactaaagtga 1635  
QY 1371 GAAAGACATGCTGATTTCTGCAGATCTCCGACATTAATGCTGTACAAATCTCGTATGG 1430  
DB 1636 catcagaagccttacccttctctgtgaggtatcccaatcagaataagattccaaag 1695  
QY 1431 ---CTATCAGGGAAGATGGAATCAATGAGTCTACTGATGATTAAGAACTACGGTTTC 1487  
DB 1696 acaatggaatctgtaattgagctcagatacagctacaataaagaagccagcgaaac 1755  
QY 1488 TTGGGCAAAAGCAAGTTTAACTGCTGCTGACGACGAGCTCCGCTATCTTCTATCT 1547  
DB 1756 ttggacaacaaagagatttctccagcccggaagaanaactcgcgtgtagtacaat 1815  
QY 1548 TCTTTGGGTTCTTTTATGATGTTCTGCTCCCTCCAAATTTATAGAGCTAGTACTGA 1607  
DB 1816 cctatgggaggtcttctactgacatctgctcgcacaacagcttgtagagtgcggaac 1875  
QY 1608 AGTGCTCTTACGAAAAAGATTTGGTTGCAAGCATTTCCAAATGTTTGCATAGAG 1667  
DB 1876 tgtatggaacacaacaaagttcttggttctccatgagaactctctgataaagac 1935  
QY 1668 CGGTGCGAATATCAAGGAATTTCCGTCATGTGAGTGAAGTGTCTATAGTGTCTAG 1727  
DB 1936 tggagatgaaatcgcgaaggtcttcgctacactcgtgagctacgtlccgttggaag 1995  
QY 1728 CAGGAGATGCGGGGTGATACCTGTCTGAGTTTGGTCAGAGCTTTGCGCGTGA 1787  
DB 1996 tgcataactcccaagaagcactatctacacttgctgcctcgaactctcttgtaga 2055  
QY 1788 CAAAGACTACTT--TATGATTAACAAATTTGCAAAAGCACTACGAGANTCTTACGTTG 1845  
DB 2056 caaagattgattatctcgtaacaacaactctagaacctcagtggaacttattctc 2115  
QY 1846 CAGCAGATGCTTCCCTATATCTGTGTGATGATATCTTTAGAGAGAGAGACATCCG 1905  
DB 2116 aagcactctatccactacaaccccaaaactatttgaattgaagaagaagaagttct 2175  
QY 1906 GAGATCCGTTGCTTATGTTTCCAAAGACTGTCGCGGTCTTCTTATGAGGAGCTTAGC 1965  
DB 2176 gaatcagctatagaaaaattccctagggaaatccctcagcttgatggtccaaagttcg 2235  
QY 1966 TACGGCATACGATCATGCAATGAAGACCGAGTCTCTTACCCCCCCCCCGAGCGTC 2025  
DB 2236 ttcagcattcagacaacgltatgaaacgactatacctcatcttgcgaagatc----- 2288  
QY 2026 TCGACGATATATCTTCTTGGGAGAGATATGCTGGGCTTGAAGAGCTGGGAACCTGAGTT 2085  
DB 2289 -----cgaaggttcttggaacagagttatagcttggttgatcggctagacct 2340  
QY 2086 GCTGTTGAANAATACGAGCGGCAAGAGATTTTCCGAGAGTACACTCATTTGTAAGTCT 2145

OY	1153	TTGGTTGCAAAACACCGAAACTATTCAGTTTAAGAAATTTGGAAATTAATATAGACGTCCTC	1212
Db	1440	--tctgaaccctgcgtctactacgaaccataacaatttgcctataataacatgaattctata	1497
OY	1213	AGGAACGAGAAAAAGATAAAATCTCACTGCTGCCACAGCTCAGAAAGATATTCGTATGAT	1272
Db	1498	gacggttcaagaagaagaaanaataagaaccacaagctacgtccaanaaatctgaattact	1557
OY	1273	CGTCTGTGTACTGCGCAATTACCGATGAGAGTTTATTCAAATAGCGCTTTTGAAATAG	1332
Db	1558	ggaaccatcacttatttgaaccgcgaagggacaagtttatgaaatacatagtttaagaat	1617
OY	1333	GACCAATTCATGATGATGGATTTCTTGATTAGATAGCTGGGAAAGACATCGTATTTCTCA	1392
Db	1618	ccctagctccctcagaacatcttagagctccaagcttccttgaactctgtatacaagccgcag	1677
OY	1393	GATTCTCGCATATAAATGCTGTACAACTCCGATAGCTATCAGGGAATGCG----	1446
Db	1678	atcccaagatccctataatggttgaagaaattccattacgctatcaaggaacttggccca	1737
OY	1447	ACAAATCAATTGGTCTACTGATGATAAGAAAGCATGCGTTTCTTGGGCAAGCAAGTTT	1506
Db	1738	atgttttgggggaaggggcttccacgaactgaaccttcacatgacccaacttggcat	1797
OY	1507	AATCCCACTGCTGCAGAGAGCGCTCCGTTAGTTTCCATCTCTTGGGGTCTTCTTATA	1566
Db	1798	atccctaaccgagcggtalcgcgctctttagtccctaagttatgtaatgcat	1857
OY	1567	GATGTTGCTCCCTCCAAATTTTATAGACATAGTACTGAAGGTGCTCTTACGAAG	1626
Db	1858	gatatcagctctccctcaattacttataaggaactgaacaaaggttgcaggaagacgt	1917
OY	1627	AGATTTTGGGTGAGGACATTTTCCATGTTTGGCATAGAGAGCGCTCGAAATTAAG	1686
Db	1918	gcttttgggtgctgattattacttacttcataagaatagtaacaaaacagagcg	1977
OY	1687	AAATTCGTCATGTGAGTGAAGTCTGTATGATGATGCTAGACAGAGATGCCGGTGCT	1746
Db	1978	gggttcgcacatttgaagtgcggttgcataagagagaaactacatactgtccaagat	2037
OY	1747	GATTCCTTGTCGCTGGGTTTGGCTGACGCTTTGGCCCGAGCAAAAGACTTTATGAT	1806
Db	2038	aagatctcttagtctcatcttgcacgctcttcttgaagagataagagactacttggatc	2097
OY	1807	ACCAATTTCCGAAGAAGCTACGCGAGAGCTTTTACGTTTGCACACAGATGCTTCCCTATAC	1866
Db	2098	aagatacaagtaacagctacgcgaaggaactctctattaccagcacaaagaaactatac	2157
OY	1867	TCTGTGCTGAGTATCCCTTTTAGAGAGAGGAGACTCCGAGATCTCTGTGCTTATATT	1926
Db	2158	tctcttcccttgc-----aaactagcgcttgcgttgcattatglt	2199
OY	1927	TCCAAAGACTGCGCTGCTCTTCTTATGGGACAGCTTACGAGCATACGGATCATGCG	1986
Db	2200	ccctacagagatctcgtctcttcttcaagaaacttagtaaccatacgaatgataaagat	2259
OY	1987	ATGAAGACGAGTCTTACCCCCCCCCCCCGAGCGCTCGAGCGATCATACTTCTTGG	2046
Db	2260	ctgaaaaacaagataaacaatacttactagttaaagaagcttgggggaatagatgttc	2319
OY	2047	GGAAGATATCTGCGGCTGGAGAGCTGGGGAACTCGATGTGCTTGAAATATCCAGCGCG	2106
Db	2320	gctttagaatctgggttgaagagctccgatttgcctagat-----gaa	2361
OY	2107	AGAGGATTTTCCGAGAGTACACTCCATTTGTMAAAGTCCAAAGCTGTTTACTCGGCGAA	2166
Db	2362	agtgctattttagcagctacacgcctccctacgaaatgcagtttgcctatgcacatag	2421
OY	2167	GATAGCTTTGTCGATCAGAGAGCTATCAGTCGAGATTTTAGGATTTGCATCTTATTAAT	2226
Db	2422	gaaggttcttaagaagaacgggaaagaagctctgtaatttggaaagtaagcgccttgat	2481
OY	2227	CTTGCAGTACCTCTTGGATCAAGTTTAGAAGAAAGGTTTGCAGACCAATAT-----TAT	2280

Db	2482	cttgccttaacctatcgggatactccgatttgataaggaatcagactgcgaagtac	2541
Oy	2281	CATGTTAGGAGATATTTCCAGATGTTTGCTGACTAAACCCAAATGTAGACTTACC	2340
Db	2542	aatcaactcttggtatactcgtgattctcttcgtagtaaccccgactgacgacaca	2601
Oy	2341	CTACTTTCACACCAAGGAGTTGGAAAGCCAAAGTTTCGACTTATAGCAAGACAGCTGGT	2400
Db	2602	ctcggaattagcgggtgattctcttgaaaacctcggtaagaatttgcagaagaacttla	2661
Oy	2401	ATTGTTCAAGGCTCAGGTGTTTCGATCTTTGGGAGCTGCAGCAGACTTTTCGGGAACCTTT	2460
Db	2662	gtccctcgtgcaggaaggaaccattttgctttaacctcaattttagagcctttagccaatt	2721
Oy	2461	GGCTTTGAAATGGCGGGATCTTTCGTGACTATATGATAGTGGCGGTAGCAAAATCAA	2520
Db	2722	tctttgatttcgtggtgcatctgcgaatcacatgtagacttagagagacaataaccaa	2781
Oy	2521	TTTTA 2525	
Db	2782	ttctta 2786	
RESULT 11			
AAA30848			
ID	AAA30848	standard; DNA; 2784 BP.	
AC	AAA30848;		
DT	29-AUG-2000	(first entry)	
XX			
DE	Chlamydia antigen CPN100634 gene open reading frame.		
XX			
KM	Chlamydia antigen; diagnosis; infection; community acquired pneumonia;		
KM	therapy; upper respiratory tract disease; bronchitis; sinusitis;		
KM	asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma		
XX	ds.		
OS	Chlamydia pneumoniae.		
XX			
PH	Key	Location/Qualifiers	
FT	CDS	1..2784	
FT		/tag= a	
FT		/transl_except= (pos:478..479; aa:Ser)	
FT		/transl_except= (pos:519..520; aa:Leu)	
FT		/transl_except= (pos:2069..2073; aa:Ile)	
FT		/partial	
FT		/product= Chlamydia antigen CPN100634	
FT		/note= "no stop codon given"	
XX			
PN	WO200032794-A2.		
XX			
PD	08-JUN-2000.		
XX			
PE	01-DEC-1999;	99WO-CA01147.	
XX			
PR	01-DEC-1998;	98US-0110339.	
PR	01-DEC-1998;	98US-0110340.	
PR	01-DEC-1998;	98US-0110427.	
PR	01-DEC-1998;	98US-0110428.	
PR	01-DEC-1998;	98US-0110438.	
XX			
PA	(CONN-)	CONNAUGHF LAB LTD.	
XX			
PI	Murdin AD, Oomen RP, Wang J;		
XX			
DR	WPI: 2000-412339/35.		
XX	P-PSDB: AAY90236.		
PT	Nucleic acids encoding polypeptide antigens from Chlamydia useful for preventing, diagnosing and treating diseases such as community acquired pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset		

OY 2047 GGAGGATATGTCGGGCGAGAGCTGGGACATCGAGTGTCTGTTGAATAATACAGCGGC 2106  
 Db 2420 gctttagaatcggcggagagcctccgattgcttagat-----gaa 2461  
 OY 2107 AGAGAGATTTTCCAGAGTACATCTCATTTGTAAGTCCAGAGCTGTTACTCGCGCAA 2166  
 Db 2462 agtgcctattatgagcagatatactgacctcctcgaattgcaattgctgctatgacacatag 2521  
 OY 2167 GATAGCTTTGTTGACAGTACAGATATCATCTGATTTTACTGATTCGATCTTTATTAAC 2226  
 Db 2522 gaaagttttaagacacagagacagacagctcgtaattggaagtagccgctctgtgaat 2581  
 OY 2227 CTTCGATCTCTCTGATGATTAAGTATAGAGAACGGTTTGAGAGCAATAT-----TAT 2280  
 Db 2582 ctgccttacctatcgagatccgatttgataagaatcagactgcgcaagatgcacgttac 2641  
 OY 2281 CATGTTGTAGCGATGATATTCATCTCATGATGTTGTGTGATGTAACCCCAATATGATACATACC 2340  
 Db 2642 aatctaactcttggttatactgctgtagctctgtagtaaccccgactgtagacacaa 2701  
 OY 2341 CTACTTTCACACAGGAGATGAGAGCAAGGTTGCACTTGAACAGACAGAGCTGGT 2400  
 Db 2702 ctgcgaattagcgggtgctcttggaacacctcgtagcgaattggaagaagacagcttta 2761  
 OY 2401 ATTGTTGAGGCTCAGGTTTTCGATCTTTGGAGCTGAGAGCTTTTGGGAACTTT 2460  
 Db 2762 gtccctcgtagcaggaaccatttctgcttaactcaatttgaagccttagccaattt 2821  
 OY 2461 GGCCTTGAATGCGGGGATCTTCTGATAGCTATATGATGATGCGGGTATGCAAAATCATA 2520  
 Db 2822 tctttagattcgctggtgctctcgcgaatttagacttagagacaaataacaa 2881  
 OY 2521 TTTTA 2525  
 Db 2882 ttctta 2886  
 RESULT 10  
 AAX06820 standard; DNA; 2787 BP.  
 AC AAX06820:  
 DT 26-APR-1999 (first entry)  
 XX Chlamydia pneumoniae surface exposed protein Omp8 DNA.  
 DE Chlamydia pneumoniae surface exposed protein Omp8 DNA.  
 XX Omp8; outer membrane protein 8; surface exposed protein; antigen;  
 KM infection; diagnosis; vaccine; atherosclerosis; asthma; ss.  
 XX Chlamydia pneumoniae.  
 OS  
 PN MO9858953-A2.  
 PD 30-DEC-1998.  
 PF 19-JUN-1998; 98MO-DK00266.  
 PR 23-JUN-1997; 97DK-0000744.  
 XX  
 PA (BIRK/) BIRKELUND S.  
 PA (CHR/) CHRISTIANSEN G.  
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;  
 XX Mygind P;  
 DR WPI: 1999-105610/09.  
 DR P-PSDB; AAW88421.  
 PT Species-specific test for identifying mammals infected with  
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for  
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding

PT these proteins  
 XX  
 PS Claim 6; Page 52-53; 115pp; English.  
 XX  
 CC This DNA sequence codes for the novel 90.0 kDa surface exposed  
 CC protein Omp8 (see AAW88421) of the human respiratory pathogen  
 CC Chlamydia pneumoniae. By generating antibodies against C.  
 CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)  
 CC was obtained which reacted with outer membrane proteins. The  
 CC antibody was used to identify the genes (see AAX06816-21) encoding  
 CC Omp4-Omp5 proteins (see AAW88417-28) in an expression library of  
 CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:  
 CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in  
 CC the other, and encode polypeptides of about 89.6-100.3 kDa and  
 CC about 56.1 kDa. The invention provides a new species specific test  
 CC for identifying mammals (including humans) infected with Chlamydia  
 CC pneumoniae. The test comprises detecting antibodies specific for  
 CC Omp4-Omp5 or detecting nucleic acid fragments encoding these outer  
 CC membrane proteins, especially by PCR. The proteins are also used  
 CC in the diagnosis of C. pneumoniae infection in mammals. The  
 CC nucleic acids and proteins can also be used in the immunization of  
 CC mammals, the nucleic acids being particularly useful as DNA  
 CC vaccines for effecting in vivo expression of antigens. The  
 CC vaccines may also prevent atherosclerosis and bronchial asthma,  
 CC which are possibly associated with C. pneumoniae.  
 XX  
 SQ Sequence 2787 BP; 811 A; 583 C; 598 G; 795 T; 0 other;  
 Query Match 13.4%; Score 338.6; DB 20; Length 2787;  
 Best Local Similarity 51.1%; Pred. No. 1.7e-88;  
 Matches 1014; Conservative 0; Mismatches 914; Indels 57; Gaps 7;  
 OY 559 AATAAGTCCCTTTCTTTTATTGGAATAGTTCTTCAACACGCTGGCGACATTCATACC 618  
 Db 841 aatcagatgcttactcttcacgaacaatacgaacagcgagagctatctatgtg 900  
 OY 619 AAAAAGCTCAGCATATCTTGTGTTGGGGAACCTATTTTCAAGGATATACAGCCCTACG 678  
 Db 901 aaaaagctcgaactggtcttcggagagactcctatcagtagaataatagtgcaatgga 960  
 OY 679 G---CTGCTGTAAGAGAGGTGCTATCGCATGCGATGCGACCCATATTCATTTCT 735  
 Db 961 gttacagctctcctaaagttgagccatagctatcgaagaatgtagtgatgattacc 1020  
 OY 736 GGACACAGTGGCGACATTTATCTTTGAAGGCAATAGATAGAGCTACAGAACCGTCTCT 795  
 Db 1021 gccgatagtggtgacatcttcttttagggatacactccttactactcctcggagacg 1080  
 OY 796 CAT---AGTCTATTGATTTAGAACTAGAGCTTAAGATTAAGCTTACGCTGCGCCAA 852  
 Db 1081 aatagaagtagtalcgactcagacagagtgcaagaatgacagcttgcgtcgcgcct 1140  
 OY 853 GGACATAGATATCTTTATGATCGATTAATCTGTAACAGAGATGACATCTGTTGCTGAT 912  
 Db 1141 gttagagccatctactctatgataccatacagatctccacaacagttacagat 1200  
 OY 913 GCTCTCATATTTATAGCCCTGATACGTGAGATTAACAAGGTTATAGGGAACTATAGTC 972  
 Db 1201 gctctaaagttaatgagactccgagatctcgaactacataatacaggaatcaltac 1260  
 OY 973 TTTTCTGAGAGAGCTCAGGAGAGCTTAAGATGAGAGTGAACCCGACCTTTAA 1032  
 Db 1261 ttcaagagagaagatcagagacagagcgagatctcctaaatcttaacttgaag 1320  
 OY 1033 TTACTCAAAATGTTGCTTTTAAATGAGATGATTTTAAAGGTGATGCTGTTTA 1092  
 Db 1321 ctactaagcctgtaactcttcagagagtagctatctttaaacaatgtagactctg 1380  
 OY 1093 AGTGGAAAGGTTTCTCTCAGAGATGCAAACTGTAAGTTGATTTAGGATGAGAGCTGC 1152  
 Db 1381 cagactcagatctcactcaacagcagagatctcgtcgaatgagcgtgagactac- 1439

PF 29-OCT-1999; 99WO-GB03579.  
XX  
PR 29-OCT-1998; 98US-0106070.  
PR 01-MAR-1999; 99US-0122066.  
PR 27-OCT-1999; 99US-0428122.  
XX  
PA (CONN-) CONNAUGHT LAB LTD.  
PI Murdin AD, Oomen RP, Dunn PL;  
XX WPI: 2000-365569/31.  
DR P-PSDB; AAY94327.  
XX  
PI Novel Chlamydia 98 kDa putative outer membrane protein antigen, used  
PI for vaccination and protection against Chlamydia infection -  
PS  
PS Claim 1; Fig 1; 93pp: English.  
XX  
XX The present sequence is the 98kDa putative outer membrane protein gene  
CC from Chlamydia pneumoniae. The genomic sequence was amplified using two  
CC PCR primers. The 5' primer contains a NotI restriction site, a ribosome  
CC binding site, an initiation codon and a sequence close to the 5' end of  
CC the 98kDa putative outer membrane protein coding sequence. The 3' primer  
CC contains the sequence encoding the C-terminal sequence of the putative  
CC outer membrane protein and a BspI restriction site. The stop codon was  
CC excluded and an additional nucleotide was inserted to obtain an in-frame  
CC C-terminal fusion with the Histidine tag. The PCR product was cloned  
CC into a eukaryotic expression vector (pCA-Myc-His) by restricting both  
CC the vector and the PCR product with NotI and BamHI and performing a  
CC ligation reaction. This expression vector was injected intramuscularly  
CC and intranasally into mice, which were subsequently inoculated with  
CC Chlamydia pneumoniae. The chlamydial lung titers of the immunised mice  
CC were lower than those of the controls. Thus the 98kDa putative outer  
CC membrane protein can be used as a vaccine to provide protection against  
CC Chlamydia infections, especially Chlamydia pneumoniae infections.  
CC The polypeptide may also be administered orally to treat Chlamydia  
CC infection. The present sequence may also be used in the  
CC construction of attenuated Chlamydia strains that can over-express the  
CC gene or express it in a non-toxic form.  
XX  
XX Sequence 3000 BP: 871 A; 617 C; 634 G; 878 T; 0 other:  
S0

Query Match 13.5%; Score 340.2; DB 21; Length 3000;  
Best Local Similarity 51.1%; Pred. No. 6.1e-89;  
Matches 1015; Conservative 0; Mismatches 913; Indels 57; Gaps 7;

QY 559 AATAAGTCCCTTTCTTTTATGGAATAGTCTTCAACACGTCGGGAGCGATTATACC 618  
DB 941 aatcagatggtactctcttcagacaacatcatcagacacgcygagagctactatg 1000  
QY 619 AAAAAGCTCACACTATCTCTGTCGGGGAACTCTATTTCAGGGGAAACAGCGCTACG 678  
DB 1001 aaaaagcgaactggtctccggaaggaactaccatcactagtaagaatagtgcaatgga 1060  
QY 679 G---CTGCTGTGAAGAGAGGTGCTATCGCATTCGAGACTGTGGACACCTTATCATCTTCT 735  
DB 1061 ggtacagctccctaaaggggagccatagctatcagaaagtaggggaattgaatllac 1120  
QY 736 GGAGCAGTGGCGGACATTATCTTTAAAGCAATACGATAGGAGCTTACAGGAACCTGCTCT 795  
DB 1121 gccagatagtggtacatgtcttctttagggaatacagctactctactactcctgagcg 1180  
QY 796 CAT---AGTGCATTTGATTAGAACTAGCGGTAGATAAAGCGCTTACGTCGCGCA 852  
DB 1181 aatagaagtagtactcgtcgaaggaagggcaagagcagagcttgcgtcgtcgtcgtc 1240  
QY 853 GGACATATGATATCTTTATGATCCGATTCGATTCATACAGATCGACATCTGTGCTGAT 912  
DB 1241 ggtagagcactactctactatcgtatcccaactacacagatcatccacaacagtagcat 1300  
QY 913 GCTTCATATTTAATAGCCCTGATCTGAGATTAACAAGAGTATACGGAACCATAGTC 972

DB 1301 gcttaaaagttaatgagactccgcagatctcgtactaacaatalacagggacatc 1360  
QY 973 TTTTCTGGAGAGCTCACCGAGGACAGAGCTTAAAGATGAGACAGACGCGCTTCTTAA 1032  
DB 1361 ttacagaggaagaagtatcagagacagagcgccagatctcctaaatctctcgcgaag 1420  
QY 1033 TTAATCAAAATGTTGCTTTTAAAGAGGAGCTGATGTTTAAAGGTGATGCTGTTTAA 1092  
DB 1421 ctactacagcctgtaactctcttcagaggtactctactatcttaaacatagagtgactc 1480  
QY 1093 AGTGGAGACGTTTCTTCAGAGATCGAAACTTAAGTTGATTTATGATTTAGGAGCTG 1152  
DB 1481 cagactcagagcactcactacacaggaagatctcgtctcgaatagagcgtgaaactac- 1539  
QY 1153 TTGCTGCAAAACACCGAAAGATTCGATTAACGATTTGGAATTAATTAACATCTCTC 1212  
DB 1540 --tcagaacctgctgataactcagccataacaaatctggtcatttaacatcgtctata 1597  
QY 1213 AGGAACGGGAAAAAGATTAACCTAGTGTGCGACAGCTCAGAAAGATTTGATAGAT 1272  
DB 1598 gacggtgcaagaagagcaaaaatagaaaccaaagctacgtcaaaaaatctgactatct 1657  
QY 1273 GGTCTGTTGTACTGCGCAATTAAGCATGAGACTTTTATCAAAATGCTTTTGAATGAG 1332  
DB 1658 ggaaccatcacttatttgaccccgacggcagcttltatgaaatacatagttlaagaaat 1717  
QY 1333 GACCATTCCTATGATGGGATTCGTGATGATGCTGGGAAAGACATCGTATTTCTGCA 1392  
DB 1718 cctcagtcctcagacatccttagagctcaaaagctctcgtacgttaacagacgcagatg 1777  
QY 1393 GATTCCTCAGATTAATGCTGTACAAATCTCCGATTCGATTCAGGAAAGTG----- 1446  
DB 1778 actcagatcctataatggtgtagaattcattcactacgctacaggaacttggggccca 1837  
QY 1447 ACAATCAATTTGGTCTACTGATATAGAAAGCTACGTTTCTTGGGCAAGCAAAAGTTT 1506  
DB 1838 atgtttggtgagacaggggtctcactgactgcaacctcaactcaactaaactaaactg 1897  
QY 1507 AATCCACGCTGAGCAGAGAGGCTCGTTAGTTCCTAATCTTGGGGTCTTTTATA 1566  
DB 1898 attcctaatcccgagcgtatcgcttcttagcctttagccttaagtatagatgacttata 1957  
QY 1567 GATGTTGCTCCCTCCAAATTTTATAGACTAGTACTGAAGGTCTCTTACGAANAAG 1626  
DB 1958 gatattaactcctcatctatctatgtagactgcaaaagaggttgaagagagacgt 2017  
QY 1627 AGATTTGGGCTTGACGACATTTCCAAATGTTTGCATAGACGGGTCTGCAAAATCAAG 1686  
DB 2018 gcttttgggtgagatcattcactcctcatalaagatagatlaacaaaacacagacgc 2077  
QY 1687 AAATTCGTCATGTAGTAGAGTCTGTAGTAGGTGTCAGCAGCAGAGATGCCGGGTGCT 1746  
DB 2078 ggggttcgcatttgaaggggtatgcatatagagaagaaactacatacttgctcagat 2137  
QY 2138 aagattcttagtgcgcaatttgcagctctttagaagagatagagactacttgtagt 2197  
QY 1807 ACCAATTTGCGAAACACTACGACAGATCTTACGTTTGACACAGAGATGCTTCCCTATAC 1866  
DB 2198 aagaatcaaggtacagctcctcactcactcactcactcactcactcactcactc 2257  
QY 1867 TCTGTGTGATGATCTTTTATAGAGAGGAGACCTCCGAGATCTCTGTGCTTATGTT 1926  
DB 2258 tctctcctctgca-----aaactaagcccttgcttgcttctatgtt 2299  
QY 1927 TCCAGACTCTGCGCTGCTTTTATAGGAGAGCTTACGATCGGCAATACGATCATGCG 1986  
DB 2300 cctacagagatctcgtctctcttcaagaaaccttaacacacatacagataacagat 2359  
QY 1987 ATGAAGACCGATCTCTACCCCCCCCCCGAGCGTCTGACAGGATCATCTCTTGG 2046  
DB 2360 ctgaaaccaaagtatacaacatactactgtttaaagaaagctggggaatagatgttcc 2419

Dp	97048	GACAGGAAACATTACCTTTGTAAGAAATACCTTACCAACACCGGAGAACGATACCTCC	96988
Oy	736	----CATAGCTGATTTGATTTAGCAACTAGCCCTAAGATACTGCGTTACGTGTCGCA	851
Dp	96988	TAAAGCTATGGGATCAACATCTAGGAAGTAAACGGAAATTCACGGAAATTCGGGCTGTAA	96928
Oy	852	AGGACATACGATATCTTTATATAGCCGATTAACGAGATGCATCTGTGGCTGA	911
Dp	96928	AAATCATACAAATTTCTTCTATATGATCCATCTTCACCTTCAGAGAACCTCAT-----CAGA	96875
Oy	912	TGCTCTCAATATTAATAGCCCTGATACCTAGGATTAACAAGATATACGAGCAACATAGT	971
Dp	96874	CGTATTTGAAGATAAATTAACGGCTGCGCGGAGCTCGCATTCATTCAGAGAAACGATTC	96815
Oy	972	CTTTTCTGGAGAGAAAGCTACGAGGACGAGCTTAAGATGAGAAACCGACTTTAA	1031
Dp	96814	ATTCTTGGAGAAACCTTACAGCAGATGAACTTAAGTTGCTGCACAAATTTAAATCTTC	96755
Oy	1032	ATTACTTCAAAATGTTGCTTTAAAAATGGGAGCTGAGTTTAAAAAGGATGTGGTTT	1091
Dp	96754	ATTCAACGACGCCAGTCTCCCTATCCGGAGAAAGTTATTGCTTACAAAAGGAGTCACTTT	96695
Oy	1092	AAATCGGAACGGTTTCTCTCAGAGATGCAAACTCTAAGTTATGATTAAGATTAGGACGTC	1151
Dp	96694	AGAGAGCAACGACCTCTCTCAGAAAGCCGGTCTCTCGTCGGGATGAGATTACGAAACGAC	96638
Oy	1152	GTTTGTGCAAAACACCGAAAGTATGAGTTAAAGAAATTTGGAATTAATATAGCTCTCT	1211
Dp	96634	ATTATCAACTACGCTGGAGAGTATTAACAATCAACAGCAACATGAGAACTCAAGTTCCTT	96575
Oy	1212	CAGGAACGGGAAAAAGATTAACCTCAGTCTGCCACAGCTCAGAAAGATATTCGATATA	1271
Dp	96574	AGGCTTTAAGACGCCCGCTAGCCTTAACACCAAAAGGTGCTTCAATTAAGTATGCTATC	96515
Oy	1272	TGCTCTGTTGACTGCGCAATTAAGGATGAGATTTTTCAAAATGGCTTTTGTGATGA	1331
Dp	96514	TGGGAAGCTCAACCGATGATTAATGAAGGAAACATTAATGAAGTCATATTTTCAGCCA	96455
Oy	1332	GGACCATTTCT-----ATGATGGGATCTTGAGTTAGATGCTGG	1370
Dp	96454	TGACACGCTCTCTCTATTAATAATCAGGTTATGCTGATGTGATTAACGTTGA	96395
Oy	1371	GAAACACATTCGATTTCTCGAGATTTTCGCGAGTAAATGCTGTACATCTCCGATGG	1430
Dp	96394	CATCAGCAGCCTTATCCCTGTTCTGCTGAGGATCCTAATTAGAAATCGGATCCCAAG	96335
Oy	1431	CTAT---CAGGGAAGTGGACACATCAATTTGGTCTACTGATGAAGAAAGCTACGCTTC	1487
Dp	96334	ACAATGGAATGTTAATTTGGAAGTACGAGTAACTACCTACAAATACAAAAGAGCCACGGCAAC	96275
Oy	1488	TTGGCAAAAGCAAGTTTAAATCCCATCTGCTGAGCAGAGAGGCTCCGTTAGTCTTAATCT	1547
Dp	96274	TTGGACCAAAACAGGATTTGTTCCAGGCCCAAGAAATCTGCGTTAGTATGCAATAC	96215
Oy	1548	TCTTTGGGTTCTTTATATAGATGTTCTCCCTTCCAAAATTTATATAGCTATGATCTGA	1607
Dp	96214	CCTATGGGAGTCTTTTACTGAAATTCGCTCTCTGCAACAGCTTGTAAGATGAGGCGCAAC	96155
Oy	1608	AGGTCTCTCTTACGAAAGAGATTTTGGGTTGACAGCAATTCATATGTTTCTCATAGAG	1667
Dp	96154	TGTTATGGAACCAACAAGGTTTCTGGGTTTCCTCCATGACGAACCTCTCTCATAGAC	96095
Oy	1728	CACGAGGATGCCGGGTGGTATACCTTGTCTGTGGGTTTGTCTCAAGCTTTTGGCGGTGA	1787
Dp	96034	TGCTCACACTCCTTAAAGACGACCTTTTACCTTTCGTTCTGCACTCTTGTGCTAGAGA	95975
Oy	1788	CAAGACACTTTATGATTAACAAATTTGGCAAAAGCACTACGAGAGATCTTACGTTTGA	1847
Dp	95974	CAAAATATGTTTATGCTCAACAACACTTGAACCTTACGGTGGAACTTTATTTCTCAA	95915

QY	1848	GGCGATGCTTCCCATATCTCTGTGGTAGTATCCTTTTAGAGAGAGGGAGCTCCGGCA	1907
Db	95914	GGACCTCTACACCTCACACCCCAAAACTATTGTAGATTAGAGACGACAAAGTTTCTGA	95855
QY	1908	GATCTGTGGCTTATGTTTTCCAGACTCTGCCGTGCTCTTCTATAGGGACCTTAGCTA	1967
Db	95854	ATCAGCTATAGAAAAATTCCCTAGAGGAATTCCTCCAGCTTGGATGTCCAAAGTTTCGT	95795
QY	1968	CGGCCATACGGATCATTCGATATAAGACGAGTCTTACCCCCCCCCCGACGCTCTC	2027
Db	95794	CAGCCTTCAGCAACCCGATATGAAACGACATATACCTATTGCCAAGATC-----	95742
QY	2028	GACGATCATATCTTCCTGGGGAGGATATGCTCGGGCTGAGAGCTGGCAACTCGAGTTC	2087
Db	95743	-----CGAAGTCTTTGGAGCAACGAGTGTATAGCTGGTATTCGGCTTAGACCTTCC	95690
QY	2088	TGTTGAAAAATACACGCGCAGAGATTTTTCCGAGAGTACACTCCATTTGTAAGTCCA	2147
Db	95689	TTTTTCTCTTCCCAACCCACATCTCTTTTCAAGACCTTATCCACAGATGAAGTCSA	95630
QY	2148	AGCTTTACTCTCGCCCAAGATAGCTTTGTTAATCTAGAGAGTATCAGTCTGATTTAG	2207
Db	95629	AATGGTTTATGATACCAAAATAGCTTCTCGAAAGCTCTAGTGAATGCCGTGTTTTAG	95570
QY	2208	TGATTCGATCTTTTAACTGCTGATTCCTTGGAAATCAAGTATAGAGAAAGGTTTC	2267
Db	95569	TATTGGAGGCTGCTTAACCTCTCGATTCCTGTGGGTGGGAAATTCGTGCAGGGGATAT	95510
QY	2268	AGAGCAATA-----TATCATGTTGTAGCAGATGATTTCTCAGATGTTTGTCTAGTAA	2321
Db	95509	CGGAAATTCCTACACCTTATGATCTCTCAGAGATTCTTGTTCGATGCTATCGTACAA	95450
QY	2322	CCCCAATGATACGATACCTCACTTCCAAACCAAGGAGTTGGAAGACCAAGTTCCGA	2381
Db	95449	TCCCAATCTACGACGACTCTGTGATGAGCCCAACACTCTTGGAAATTCGCGGTGGCA	95390
QY	2382	CTTAGCAGACAGCGCTGATATGTTCAGGCCCTCAGAGTTTTCGATCTTTGGAGCTGCAC	2441
Db	95389	TCTTTCAAGACAGGAGATTTTACTAGGGGTGCAAAACATCGCTACAACTCCAAATTG	95330
QY	2442	AGAGCTTTTCGGGAACCTTGGCTTTGGAATGCGGGGATCTTCTGATATATGTAGA	2501
Db	95329	TGAGCTCTTCGACATTTACGCTATGGAACCTCGTGGATCTTCAAGAACTACAAATGTAG	95270
QY	2502	TGCGGGTAGCAAAATCAATTTTAG	2526
Db	95269	TGTTGTTACCAACTCCGATTTCTAG	95245
RESULT 9			
AA	AA27021		
ID	AA27021	standard; DNA; 3000 BP.	
AC	AA27021;		
XX	11-AUG-2000	(first entry)	
XX	Chlamydia pneumoniae	98kD putative outer membrane protein gene.	
XX	Chlamydia; antigen; vaccine; infection; outer membrane protein; ds.		
XX	Chlamydia pneumoniae.		
XX	OS		
XX	Key	Location/Qualifiers	
XX	CDS	101..2887	
XX	FT	/*tag= a	
XX	FT	/product= "98kDa putative outer membrane protein"	
XX	FN	MO200026237-A2.	
XX	DB	11-MAY-2000.	
XX			







```

OY 681 TCGTGTAAAGAGGCTGCTATCGCATGTCAGACTCTGACACCCCTATCCATTCTGAGAGA 740
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1024 aactcctaagg999gtgctatcagcatcgatgctccagagagccagctctcttcgcgaga 1083
OY 741 CAGTGGCGACATTTATCTTTGAAGGCAATACGATAGAGCTTACAGAACCCCTCTCT- 795
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1084 gacaggaacattaccccttgtaagaatacccttaacaacaggaagtlacgcgatactcc 1143
OY 796 ----CATAGTGTATTTAGTTAGAACTAGAGCTTAAGTAACCTGGCTTACGTGCTGCCA 851
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1144 taacagctaatcgatcaacaataaggaagaaacggaataacggaataacggaatcgctaa 1203
OY 852 AGGACATCGATATATCTTTATGATCCGATTACTGTAACAGGATGAGATCTGTTGCTGA 911
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1204 aatcatacaattctctctatgataccatccatccagaaagaaactcat-----caga 1257
OY 912 TGCTCTCAATTAATTAAGCCCTGATACCTGAGATTAACAAGATTAACGGGACCATAGT 971
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1258 cgtattgaagataaataacgctcctcgagagctctcaatccatacgaagacgactct 1317
OY 972 CTTTCTGAGAGAAAGCTACGAGAGCAGAGCTTAAGATGAGAAACCGCATCTTAA 1031
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1318 attctctggagaaaccccaacagcagatgaactaaagttgctgcaatlttaaatcttc 1377
OY 1032 ATTACTTCAAAATGTTGCTTTTAAATGGGACTGTAGTTTAAAGGTGATGCTGTTT 1091
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1378 attcagcagcagctccctccatcccgaggaagtattctgcaacaagaagagctactt 1437
OY 1092 AAGTCCGAAGCGTTTCTCTCAGAGATGCAAACTCTAAGTTGATTATGCAATTAAGCACTG 1151
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1438 agagagcagagctctctcaagaagccggtctctcctcgagcagatctcagcagac 1497
OY 1152 GTTGGTTCGAAACACCGAAGATGATGAGTTAAAGTAATTAATATATGATCTCT 1211
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1498 attatcaactacagcttggagatatacaatacgaacctaggaatacaatgttgcactt 1557
OY 1212 CAGGAACGGGAAAAAGATAAATCTCAGTCTGCCACAGCTCAGAAAGATATTCGTATGA 1271
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1558 aggtcttaagcagcccgctcagccttaacagcaaaagtgcttcaataaagtgatcgatc 1617
OY 1272 TCGTCTCTGTTACTGCGAATTAAGCATGAGAGTTTATCAAAATGGCTTTTGAATGA 1331
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1618 tgggaagctcaacgtatgataatgaaagacattatgaagctatglatgacga 1677
OY 1332 GGACCATTCCT-----ATGATGGATTTCTGATTAGATGCTGG 1370
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1678 tgacagactctctctcattaaatacaacggttgaatgctgactgatacactaacttga 1737
OY 1371 GAAGACATCGTATTTCTGCGAATTCGCGAGTAAATGCTGTACATCTCCGTATGG 1430
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1738 catcagcagccttaccctgctcgtcgtgagatcctaattcagaatacgcgttcaag 1797
OY 1431 CTAT---CAGGGAAGTGGCAATCAATTTGGTCTACTGATGATTAAGAAAGCTACGCTTTC 1487
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1798 acaatggaatgttaattgactacgatacagctcaatacaagaagccacgagcgaac 1857
OY 1488 TTGGCAAGAGCAAAATTTTAATCCACTGCTGAGCAGAGGCTCGCTAGTCTTAATCT 1547
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1858 ttggaccacaacaggaattgttccacagcccgaaagaatactcgcttaagtatac 1917
OY 1548 TCTTTGGGTTCTTTTATAGATGTGCTCCCTTCCAAAATTTTATAGACTAGTACTGA 1607
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1918 cctatgggagcttcttactgacatcgtctctgacacagcttgaagatcgcgcaac 1977
OY 1608 AGTGTCTCTTAACGAAAGAAATTTGGGTTGCGAGCATTTTCCAATGTTTTCATAGAGAG 1667
   || || || || || || || || || || || || || || || || || || || || || ||
Db 1978 tggattggaacacaacaggttctcgttccctcactgacgaactcctcgcaataag 2037
OY 1668 CGGTGCTGAATTAACAAAGAAATTCGTCATGTAGTGAAGTGTGTAGTAGAGGTAG 1727
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2038 tggagatgaaaaatcgcaaaagcttccgctcaatacctcggagagctagcagtgag 2097

```

```

OY 1728 CACGAGATGCCGGGTGATGATACCTGTCTCTGGGTTTGGCTCAGCTTTTGCGCGTGA 1787
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2098 tgcctacactccttaagaagacattatcacttctgcttgcctccactcttgcagaga 2157
OY 1788 CAAAGACTACTTTATGAAATACCAATTTTGGCAAGACCTACGAGATCTTTACGTTTGA 1847
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2158 caaagattgtttatcgcctacaacaactcagaactcaggtggaacttattcttcaa 2217
OY 1848 GCAGATGCTTCCCTATPACTCTGTTGATGATCTCTTTTGAAGAGAGGGAGCATCCGGGA 1907
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2218 gcaactcataaccctacaacccccaaactatttgatagataggaagagaagtttctga 2277
OY 1908 GATCCTGTGCTTATGTTTCCAGACTCTGCCGTGCTCTTTCTATGGGACATTAGCTA 1967
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2278 atcagctatagaaaaattcccttaggaattcccttagacttgatgctgcaagttcgt 2337
OY 1968 CGGCCATACGGATCATTCGCATTAACACGAGTCTTACCCCCCCCCCCCCAGCGCTCTC 2027
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2338 cagccattcagacaacgctatgaaacgactatacctcatctgccaagatc----- 2388
OY 2028 GACGATCATACTTCTTGGGAGGATATGCTGGGCTGGAGAGCTGGAACTCGAGTTGC 2087
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2389 -----cgaaagttcttggagacaagagtgatagctggtgtatcgcgactagacttc 2442
OY 2088 TGTGAAAATACGACGCGCAGAGATTTTTCGAGAGTACACTCCATTTGTAAGTCCA 2147
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2443 ttgttcttccacacacacatccctcttccaagaccttccatccaagatgaagtcga 2502
OY 2148 AGCTGTACTCTGCGCCCAAGATGCTTTGTTGACTAGAGACTATCAGTCGATGTTTGA 2207
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2503 aatggttattgatacaaaaatagcttctcgaagctctagtgatgacggtggtttag 2562
OY 2208 TGATTCGATCTTTTAACTTCGATTCCTGATTCCTGATTAACATCAAGTAAGAAAGCTTGC 2267
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2563 tatggaaggtcgtctaacctctcgtatccgttgcgttgggtggaatccgagcggtgatat 2622
OY 2268 AGAGCAATA-----TATCATGTTGTAGCGATGATTTCTCCAGATGTTTGTGTATGA 2321
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2623 cggagattccctacactatgactctcagatcttcttgcgttgcgtatcaacaa 2682
OY 2322 CCCCAATGTACGATCACTCTTCCACCAAGGAGGAGTTGGAAGACCAAGTTCCGA 2381
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2683 tccccaactcagcagactcttgatgagccagactcttgaataatcgcgtgagaa 2742
OY 2382 CTATACAGACAGCGTGTGATGTTTCAGGCTTCAGGTTTTCATCTTTGGAGCTGCAGC 2441
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2743 tcttcaagacagggatlttctactgaaggtgtagcaacaactacgtctacaactcaatg 2802
OY 2442 AGAGCTTTTCGGGAACCTTTGGCTTGAATGCGGGGATCTTCTGTACTATATATGAGA 2501
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2803 tgaagcttcgacattacgctatggaactccgtgactcttcaagaactacaatgtaga 2862
OY 2502 TGCGGCTAGCAAAATCAATTTTGA 2526
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2863 tgtgtaccacaactcgaattcag 2887

```

## RESULT 7

AA06816 standard; DNA; 3200 BP.

AA06816:

26-APR-1999 (first entry)

Chlamydia pneumoniae surface exposed protein Omp4 DNA.

Omp4; outer membrane protein 4; surface exposed protein; antigen;

infection; diagnosis; vaccine; atherosclerosis; asthma; ss.

Chlamydia pneumoniae.

Key Location/Qualifiers



OY 1980 TCATGCATGAGACGAGTCTCTACCCCCCCCCCCCCGAGCTCTCGACGATCATAC 2039  
|||||  
DB 30929 tcacgcataagacagcgatctctaccccccccccccgacgctctcgacgatacatca 30988  
OY 2040 TTTCTGGGAGATATGCTGGGCTGGAGAGCTGGGAAGTGCAGTGTGTAATAATAC 2099  
|||||  
DB 30989 tctctgggagagataltgctggcggagagctgggaactcgaagtcggtgtaataatc 31048  
OY 2100 CAGCGGACGAGATTTTCCGAGATACATCCATTTTGAAGAAGCCAAAGCTTTTACTC 2159  
|||||  
DB 31049 cagcgccagagagatlttcccaagagataccatcatttgaagaagccaaagcgtttaagc 31108  
OY 2160 GCGCCAGATACCTTTGTAAGTGAAGCTAATGAGTGTGATTTTGAATTCGATCT 2219  
|||||  
DB 31109 tcgcaagaagtagctttagaactagagctacagtcgtagttagttagtgcacatc 31168  
OY 2220 TTATATACCTTGATCTCTCTGGAATCAAGTATAGAAACGCTTTCGACAGCAATATTA 2279  
|||||  
DB 31169 ttataacctggaatcctccttggaatcaagtagaagaacggttgcagagcaatata 31228  
OY 2280 TCATGTGTAGGATGTATTCCTCAGATGTTCGTAGTAAACCCAAATGATGACATAC 2339  
|||||  
DB 31229 tcagtgttagagatgtagtattcccaagatgtagttagtaaccccaaatgtaagactac 31288  
OY 2340 CCTACTTCCAAACGAGGAGTGTGAAGACCAAGGTTGGAATTCGAACTTAGCAAGACGCTGG 2399  
|||||  
DB 31289 cctacttcccaacgaagagagtgtagaagcaaaagttcgaacttagcaagcagagcttg 31348  
OY 2400 TATTTGTACGCGCTCAGGTTTTCGATCTTTGGAGCTGACAGAGCTTTTGGGAACCT 2459  
|||||  
DB 31349 tatgttcagagcctcaggttcttcgaatccttggaagctcgaagagagcttctcggaact 31408  
OY 2460 TGGCTTTGATGGGCGGAGTCTTCCTGATAGTATATGATGAGGCGGAGCAAAATCA 2519  
|||||  
DB 31409 tggccttggaagtgaggagatcttctcgtagcataatgtagtgcggtagcaaatcaa 31468  
OY 2520 ATTTTAA 2526  
|||||  
DB 31469 attttag 31475

RESULT 5  
AAK06830  
ID AAK06830 standard: DNA; 1830 BP.  
AC AAK06830;  
XX  
XX 26-APR-1999 (first entry)  
DE Chlamydia pneumoniae surface exposed protein Omp7 DNA.  
XX  
XX Omp7: outer membrane protein 7; surface exposed protein; antigen;  
KW infection; diagnosis; vaccine; atherosclerosis; asthma; ss.  
XX  
OS Chlamydia pneumoniae.  
XX  
XX MO9858953-A2.  
PN  
XX 30-DEC-1998.  
PD  
XX 19-JUN-1998; 98WO-DK00266.  
PF  
XX 23-JUN-1997; 97DK-0000744.  
PR  
XX  
XX (BIRK/) BIRKELUND S.  
PA (CHR1/) CHRISTIANSEN G.  
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;  
PI Myland P;  
XX  
XX MPI: 1999-105610/09.  
DR P-PSDB: AAM86431.  
XX

PT Species-specific test for identifying mammals infected with  
PT Chlamydia pneumoniae - comprises detecting antibodies specific for  
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding  
PT these proteins  
PS Disclosure: Page 83-86; 115pp; English.  
PS  
XX  
XX This DNA sequence encodes the novel surface exposed protein Omp7  
CC (see AAM88431) of Chlamydia pneumoniae, a human respiratory pathogen.  
CC It is described as a subsequence of a claimed nucleic acid fragment  
CC (see AAK06819) encoding Omp7 (see AAM88420). The invention provides a  
CC new species specific test for identifying mammals (including  
CC humans) infected with C. pneumoniae. The test comprises detecting  
CC antibodies specific for surface exposed proteins Omp4-Omp15 (see  
CC AAM88417-28) or detecting nucleic acid fragments encoding them (see  
CC AAK06816-27), especially by PCR. The proteins are also used in the  
CC diagnosis of C. pneumoniae infection in mammals. The nucleic acids  
CC and proteins can also be used in the immunization of mammals, the  
CC nucleic acids being particularly useful as DNA vaccines for  
CC effecting in vivo expression of antigens. The vaccines may also  
CC prevent atherosclerosis and bronchial asthma, which are possibly  
CC associated with C. pneumoniae.  
XX  
SQ Sequence 1830 BP; 485 A; 357 C; 443 G; 545 T; 0 other:  
  
Query Match 56.4%; Score 1425.2; DB 20; Length 1830;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1430; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
  
OY 873 TGATCGATTACTGTACAGATGACATCTGTGCTGATGCTCAATATTAATAGCCC 932  
|||||  
DB 393 tgatcgcattactgtaacagagatcgacatctgtgtgagctccaataataatagccc 452  
OY 933 TGATCTGAGATACCAAGAGATATACGGGAACCATAGCTTTTTCGAGAGAAACTGAC 992  
|||||  
DB 453 tgatctgagatacaagagatlaacggaacacatagcttcttctgagagaaactcac 512  
OY 993 GGAGGACAGACCTAAAGATGAGAGAACCGCACTTCTAATTAATCTCAAAAGTTGCTTT 1052  
|||||  
DB 513 ggaggcagaagctaaagatgagaagaacccgacctctaataactcaaaagtgtcctt 572  
OY 1053 TAAAAATGGGACTGTAGTTTAAAGTGATGTCTGTTTAAGTCGAAACGTTTCTCA 1112  
|||||  
DB 573 taaaatgggactgtagttttaaagtgatgtcgtttaaagtgcaaacggttctctca 632  
OY 1113 GATGCAAACTTAAGTGTATGATTTAGGACGTCGTTGTTGCAAAACGGAAG 1172  
|||||  
DB 633 gatgcaaaacttaagtgatgatttaggacgtctgtgtgcaaacccgaag 692  
OY 1173 TATCGATTAAAGAAATTTGAATTAATATGACCTCTCAGAACGCGAAAGATATA 1232  
|||||  
DB 693 tatcgattaaagaaatttggaaattcaataagactcttcaggaacggaagaaagataaa 752  
OY 1233 ACTGAGCTGCGACAGCTCAGAAAGATATTCGTATGATCGTCTGTTGATCTGGCAAT 1292  
|||||  
DB 753 actgagctgcgacagctcagaagatactcgtataagtcgtctgttactgcaat 812  
OY 1293 TAGCATGAGATTTTNTCAAAATGCGTTTGAATGAGGACCAATTCATAGGGGAT 1352  
|||||  
DB 813 tagcatgagatlttltcaaaatgaccttltgatatggagccattccatagatggat 872  
OY 1353 TCTTGAGTTAGATGCTGGGAAGACATGCTGATTTCTGAGATTTCTCGAGTATAATGC 1412  
|||||  
DB 873 tcttgagttagatgctgggaagacatgctgatttctgagagttctcgagatagatgc 932  
OY 1413 TGTACAAATCTCCGTATGCGTATCAGGAAAGTGAACATCAATTTGCTACTGATGATTA 1472  
|||||  
DB 933 tgtacaatctccgtatgctatcagggaaagtgaacatcaatttggctactcagatataa 992  
OY 1473 GAAAGCTAGGTTTCTGGGCAAGCAAAAGTTTAAATCCCAATCCGATGAGCAGAGCTCC 1532  
|||||  
DB 993 gaaagctaggttcttcttgaggcaagcagatlttaatccactgctgagcagagctcc 1052

CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
CC encoded by the open reading frames of the C. pneumoniae genome (see  
CC AY34584-Y35879) can be used in immunogenic compositions as vaccines.  
CC Vectors containing C. pneumoniae nucleotide sequences can also be  
CC used as immunogenic compositions, especially where the vector directs  
CC the expression of a neutralising epitope of C. pneumoniae.

XX Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match 98.4%; Score 2485.4; DB 20; Length 1230025;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 2514; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

QY 1 ATGAGATTCACCTCCGCTTTTATGATATCATATTAGCTACGCTTTCATGTGCAAT 60  
DB 28950 atgaagcttcacccgcttcttattgatactagctacgaagcttcctcgtatgcaat 29009  
QY 61 TTATTAGAGCTGCTACTACGAGAGCTATCGGCTAGCAATAGCTCGAATGGAATGACA 120  
DB 29010 ttattagagctgctactacccgaagatctcgtctagcatagcttcgtagaactaca 29069  
QY 121 TCACAAACAAGCTTTCTAGTAAACATCATCGGCTACAGATGGCACCAATTATGTTTT 180  
DB 29070 tcaacaacaagcttctctagtaaaacatcatcgctacagatggcacaattatgctttt 29129  
QY 181 AAGATCTCTAGTATATAGAAATGTACCCAAAACAGGGAACCTCATAGTGTGT 240  
DB 29130 aagatctctagatataagaaatgtacccaaaacagg9gaactcagctcactaagtggt 29189  
QY 241 TTTAAAAATGACGCTGCAGCTGAGATCTAAATTTCTAGAGAGGGGATTTCTTTCACA 300  
DB 29190 tttaaaaaatgacgtgcagctgagatcttaattctctagagagg9gattctcttaca 29249  
QY 301 TTTAGCAATTCGATGACACACGCGCTTTCGAGCTCTATTGGAAGTGAACGCTAAT 360  
DB 29250 ttttagcaatctcgatgaacacacgcttctcgagctctattggaagtgaagcagctaat 29309  
QY 361 AAGCAGCTCAGTTATAGATTTTGGCAGCTTCTTTTAAATCCGACGACGATACA 420  
DB 29310 aagcagctcagttatagatTTTGGCAGCTTCTTTTAAATCCGACGACGATACA 29369  
QY 421 GTGCAATGATGATGAGAGTATCAATGTTAAAGGAATTTAAAGCTTATGATATGAT 480  
DB 29370 gtgcaatgattgatgagatcaatgattaaaggaatTTAAAGCTattgataatgat 29429  
QY 481 AAGGTATGATTCAGACAAATTTCTACACAGAGATGGCGAGCAATTAATGTGACGC 540  
DB 29430 aaggtattgattcagacaaattctcacaagagatgycgga-caattaatltygcagc 29488  
QY 541 TCCCTGAAGATCCGAACATTAAGTCCCTTCTTTTATTTGGAATAGTCTTCAACAGT 600  
DB 29489 tccctgaagatccgaacataaagcttctcttcttatttgaataagttcttcaacagct 29548  
QY 601 GGGGAGAGCGATTATACCAAAAACCTCACACTATCTCTGGGGAACCTTATTTTCAG 660  
DB 29549 ggggagagcgattataccaaaaacctcacactatctctgtygg9gaactctattcag 29608  
QY 661 GGGAAATACAGCGCTACGCGTGTGTAAAGAGGTGCTATTCGATTTGACAGCTGCG 720  
DB 29609 gggaaatacagcgctcagcgctgtgtaag9gagtcacgcatltygaagctcctgc 29668  
QY 721 ACCCTTCAATTTCTGAGACATGGCGANATATCTTTTAAAGCAATACGATAGAGCT 780  
DB 29669 acccttcaatttctgagacatggcganattctttttaaagcaatACGATAGAGCT 29728  
QY 781 ACAGAGACCGTCTCTATAGTGTATTGATTAGAACTAGCCTTAAGATTAACCTGCTTA 840  
DB 29729 acagagaccgctctctatagtgctattttagaaactagcgtctaaagatactgctta 29788  
QY 841 CGTGTCGCGAAGACATACGATATCTTTTATGATTCGATTAAGTACGAGATGACACA 900  
DB 29789 cgtgctcgcgaaagacatactattttagatccggttactctgtaaaagatgcgaca 29848

QY 901 TCTGTGATGATGCTCTCAATATTTATAGCCCTGATACCTGGAGATTAACAAAGATATAC 960  
DB 29849 tctgtgatgattgctctcaatattttatagccctgatactctggagatatacaagatatac 29908  
QY 961 GGAACCAATAGCTTTCTTTCGAGAGACCTCACGAGGACAGAGCTAAAGATGACAAAGAC 1020  
DB 29909 ggaaccatagcttcttcttctgagagaagctcacgag9gcagaa9gataagatgagaagac 29968  
QY 1021 CGCAGCTTCAATTAATTAATGTTGCTTTTAAAAATGGAGCTGATGTTT-AAAAAG 1079  
DB 29969 cgcacttcaattacttcaaaatgcttctttaaataatg9gactgtagtttga9aaag 30028  
QY 1080 TGATGCTGTTTAAAGTGGCAACGCTTCTCTCGAGATGCAAACTCTAAGTTGATTAAGA 1139  
DB 30029 tgaatgctgTTTAAAGTGGCAACGCTTCTCTCGAGATGCAAACTCTAAGTTGATTAAGA 30088  
QY 1140 TTTAGGAGCTGCTGTTGGTTCACAAACACGAAAGATGCAATTAAGCAATTTGCAATTA 1199  
DB 30089 tttaggagctgctgTTGGTTCACAAACACGAAAGATgatacgatgataagcttggaaatla 30148  
QY 1200 TATAGACTCTCTCAGAGACGGGAAAAAGATTAACCTCATGCTGCTCCACAGCTCAGAAA 1259  
DB 30149 tatagactctctcagagacgg9aaagataaaactcagtgccacagctcagaaaga 30208  
QY 1260 TATTCGATATGATGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1319  
DB 30209 tattcgatatagaltcgtctcgttctgtaacgcaatlaag9aag9gttcttatacaaaatg 30268  
QY 1320 CTTTTTGAATGAGACCACTTCATGATGGGATTTCTTGAATGATGATGCTGGAAGACAT 1379  
DB 30269 ctttttgaatgagaccattcctatgattg9gattcttgaatgattgctg9aaagcat 30328  
QY 1380 CGTATTTTCGACAGATTCCTGGCAGTATAAATGCTGTATCATCTCCGATATGCTATCAGG 1439  
DB 30329 cgtatcttcgacagattctcgcagtaga9atgctgtaacatctcgtatg9ctacag9g 30388  
QY 1440 AAGTGACATCAATATGCTACTGATGAATGAAGAAAGTACGGTTCTTGGGCAAGCA 1499  
DB 30389 aagtgacatacaatattgctactgata9aaagcaag9gttcttctg9g9aaagca 30448  
QY 1500 AAGTTTAAATCCACTGCTGAGACGAGGCTCGGTAGTTCTAATCTTTTGGGGTTC 1559  
DB 30449 agtTTTAAATCCACTgctgagag9gctcgttagttccctaactctcttctg9g9ctc 30508  
QY 1560 TTTTATGATGCTTCCTCCAAATTTTATAGAGCTAGTACTGAAAGTCTCTTA 1619  
DB 30509 ttttatagatgcttcctccaaattttatagagctagta9gctag9atg9gtctccta 30568  
QY 1620 CGAAAGAGATTTTGGGTGACAGGCAATTTCCANATGTTTGCATAGAGCGGTGTAATA 1679  
DB 30569 cgaagaagatttgggtgacaggcaatttccaanatgTTTGCATAGAGCGGTGTAATA 30628  
QY 1680 TCAAGGAAATTCGTCATGTAGTGAAGTGTGTAGTAGTGTGACGAGAGATGCC 1739  
DB 30629 tcaaggaaatttcgctcatgtgtagtgaagtgtgtatgtagtagtgcagac9gagatgcc 30688  
QY 1740 GGGTGTGATACCTGTCTGTGGGTTTGGCTACAGCTTTTGGCGTGACAAAGACTACTT 1799  
DB 30689 ggggtgtgatacctgtctgtgggtttggctacagctTTTGGCGTGACAAAGACTACTT 30748  
QY 1800 TATGATATACCAATTTGCAAGACCTACGAGATGTTTACGTTTGCAGACAGATGCTTC 1859  
DB 30749 tatgatataccaatttgcagaagaccta9gaagatctttagtctgcaagac9gagatcttc 30808  
QY 1860 CCTATACCTGTGTGATATCTTTTATGAGAGGAGAGACTCCGAGATTCCTGTTGCC 1919  
DB 30809 cctatactcgtgtgtagtatactttagagag9gag9gtcgcgagatcctgtgtgcc 30868  
QY 1920 TTAATTTCCAAAGACTCTGCGGTCTTTCTATGGGACCTTAAGCTACGCGCATACGCA 1979  
DB 30869 ttatgttccaagactcgtgcgtctcttctata9g9cagcttagctacg9catatag9a 30928



Db 1061 ggaaccatagatcttcttcgagagaagctcaacgagcagaagctaagaatgagaagac 1120  
 Oy 1021 CGCACTTCCTAAATTAATCTCAAAATGTTGCTTTTAAATAGGAGCTGTATTTAAAGGT 1080  
 Db 1121 cgcactcttaaatcttcaaaaatgcttctttaaataatggaacgctgattttaaagct 1180  
 Oy 1081 GATGTCGTTTTAAAGTGCAGACGCGTTTCTCTCAGATGCAAACTTAAGTTGATTTGAT 1140  
 Db 1181 gatgcgtttaaagtcgagaaagcttctctcaagatgcgaactctaaagttatgtagat 1240  
 Oy 1141 TTAGGAGAGCTGCTTGTTGCAACACCCAAAGTATCGAGTTAGCAATTTGGAAATTAAT 1200  
 Db 1241 ttaggaagcgtcgttgctgcaaacacgaaagtaacgagttacgaatctggaataat 1300  
 Oy 1201 ATAGCTCTCTCAGAGAGGGGAAAAGATPAAACCTCAGTCGCGACACGCTCAGAAAGAT 1260  
 Db 1301 atagactctcctcaaggaaagggaaaagataaactcaatgcagtcgacacagctcagaagat 1360  
 Oy 1261 ATTGCTATAGATCGTCCTGTTGTACTGTGCAATTAAGCATGAGAGTTTATCAAAATGAGC 1320  
 Db 1361 atctgtatagatcgcctcgttgtagtactgcaatlaagcagatgagagtttatacaaaatgac 1420  
 Oy 1321 TTTTGAATGAGAGCACTTCTATGATGGGATTTGAGTTAGATGCTGGAAAGACATC 1380  
 Db 1421 ttttgaatgagacacatctcctatgtagtactgtagtactgtagtactgtagtactgtagt 1480  
 Oy 1381 GTGATTTGCGAGATTCGCGAGTAAATGCTTAAATGCTTAAATGCTTAAATGCTTAAATG 1440  
 Db 1481 gtagtcttcgcagatctcgcagatagatagatgctgacaaatcccgtaagcctacagga 1540  
 Oy 1441 AAGTGACATCAATTTGGTCTACTGATGATGAAGAAAGCTACGCGTTCTTGGGCAAGCA 1500  
 Db 1541 aagtgacatcaatcttgcttactgtagatagaagaagctacggttcttggcgaagcag 1600  
 Oy 1501 AGTTTATCCACACGCTGAGAGAGAGGCTCGTTAGTTCTTAATCTTGGGCTTCT 1560  
 Db 1601 agtttaatccacacgctgagcaagagagctcctgtagtactccttcttggggttct 1660  
 Oy 1561 TTTATAGATGTTGCTGCTCCCTCCAAATTTTATAGAGTACGAGTACGAGTACGAGTAC 1620  
 Db 1661 tttaagatgcttcttcttctcctcagaatttataagagctaggtactggaaggtgctcctac 1720  
 Oy 1621 GAAAAGATTTGGGTTGACAGCAATTTCCAAATGTTTGCATAGAGAGCGGCTGTAAT 1680  
 Db 1721 gaaaagagatttgggttgcagcatttccaatgcttgcagtagagagcgtcgtgtaaat 1780  
 Oy 1681 CAAAGAAATTCGCTCATGTGAGTGAGTGAGTGCTGTATGATGATGATGATGATGATG 1740  
 Db 1781 caaagaaatctcgtcattgtagttagttagttagttagttagttagttagttagttagt 1840  
 Oy 1741 GGTGGTATACCTTGTCTGCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800  
 Db 1841 ggtggtataccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1900  
 Oy 1801 ATGAATACCAATTTGCAAGACCTACGAGAGATCTTAAAGTTTGCAGACAGATGCTTCC 1860  
 Db 1901 atgaataccaatcttgcagaagacctaagcagatctttagcgttgcagagatcttcc 1960  
 Oy 1861 CTATACCTGCTGCTGATGATCTTTTATAGAGAGAGAGATCTCCGAGATCTCTGTTGCT 1920  
 Db 1961 ctatacctcgtgtagttagttagttagttagttagttagttagttagttagttagttagt 2020  
 Oy 1921 TATGTTTCCAAAGTCTGCTGCTGCTTCTTATGAGGAGAGTACGAGTACGAGTACGAGT 1980  
 Db 2021 tatgtttccaaagacctcgtcgtcttcttcttcttcttcttcttcttcttcttcttctt 2080  
 Oy 1981 CATGCGATGAGACGAGTCTCTACCCCCCCCCCGAGCGCTCTCGAGGATCATACT 2040  
 Db 2081 catcgatgaagacgagctcttacccccccccccccgagcgtctcgcagatataat 2140  
 Oy 2041 TCTTGGGAGGATGATGCTGGGCTGGAGAGTGGAGAACTGAGTGGTGTAAATATCC 2100  
 Db 2141 tcttgggagatagctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 2200

Oy 2101 AGCGCAGAGATTTTTCGAGAGTACACTCCATTTGTAAAGTCAAGCTGTTACTCG 2160  
 Db 2201 agcgcgagaggttcttccaagagtaacatcttcttgaagaagctcaagctttagct 2260  
 Oy 2161 CGCCAGATGATGTTGTTGACTAGAGATATCATCGTATTTAGTATTTGATTTGATCTT 2220  
 Db 2261 cgccaagatagctttagtaactagagctacacgctcgtgatttagttagttagttagt 2320  
 Oy 2221 TATACCTTGCAATCTCTTGTGATTCAGATGTTGTTGTTGTTGTTGTTGTTGTTGTT 2280  
 Db 2321 tataaccttgatctccttcttgaatcaagtagaagaacggttcgagagaataatctat 2380  
 Oy 2281 CATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340  
 Db 2381 catgtttagtagatgtagttagttagttagttagttagttagttagttagttagttagttag 2440  
 Oy 2341 CTACTTCCAAACCAAGAGGATTTGGAAGACCAAGTTGCAACTTATGCAAGACAGCTGCT 2400  
 Db 2441 ctacttccaaccaaaggaggttggagaccaaaggttcgaacttagcaagacagctgct 2500  
 Oy 2401 ATTGTTCAAGCTCAGGTTTGTGATCTTTGGAGCTGCGACGAGCTTTTGGGAATCTT 2460  
 Db 2501 attgtcagcctcaggttctgacttcttggagctgcaagcagccttctcggaaact 2560  
 Oy 2461 GCGTTGAAATGCGGGGATCTCTGCTAGCTTAAATGATGATGATGATGATGATGATGAT 2520  
 Db 2561 ggccttgaatgagggagcttcttcttcttcttcttcttcttcttcttcttcttcttct 2620  
 Oy 2521 TTTTAG 2526  
 Db 2621 ttttag 2626

RESULT 4  
 AAX91990  
 ID AAX91990 standard: DNA; 1230025 BP.  
 AC AAX91990;  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; ss.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO9927105-A2.  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-IB01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 PA (GENSET ) GENSET.  
 XX  
 PI Griffiths R;  
 DR WPI: 1999-357842/30.  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 PS Claim 1; Page 291-611; 1912pp; English.  
 CC The present sequence represents the complete genome of Chlamydia  
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes  
 CC respiratory disease such as pneumonia and bronchitis and is thought  
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,



OY 2461 GGCCTTGAATGGCGGAGTCTCTGATAGTATATAGATGCGGGATGCAAAATCAAA 2520  
|||||  
Db 2461 ggcttgaatggcgaggatctctcgtacgataatgtagtcggtagcaaatcaaa 2520  
OY 2521 TTTTATG 2526  
|||||  
Db 2521 tttag 2526  
RESULT 3  
AAA28690  
ID AAA28690 standard; DNA; 2750 BP.  
AC AAA28690;  
XX  
XX 29-AUG-2000 (first entry)  
DE C. pneumoniae CPN100397 gene.  
XX  
KM Antigen: anti-inflammatory; respiratory; antibacterial; anti-asthmatic;  
KM anti-arteriosclerotic; vaccine; ds.  
OS Chlamydia pneumoniae.  
XX  
FH Key location/Qualifiers  
FT CDS 101..2626  
FT /\*tag= a  
XX  
XX MO200024765-A2.  
PN 04-MAY-2000.  
PD  
XX 28-OCT-1999; 99MO-CA00992.  
XX  
XX 28-OCT-1998; 98US-0106034.  
PR 28-OCT-1998; 98US-0106039.  
PR 28-OCT-1998; 98US-0106042.  
PR 28-OCT-1998; 98US-0106044.  
PR 28-OCT-1998; 98US-0106072.  
PR 28-OCT-1998; 98US-0106073.  
PR 28-OCT-1998; 98US-0106074.  
PR 28-OCT-1998; 98US-0106087.  
PR 02-NOV-1998; 98US-0106587.  
PR 02-NOV-1998; 98US-0106588.  
PR 02-NOV-1998; 98US-0107034.  
PR 02-NOV-1998; 98US-0107035.  
XX  
XX (CONN-) CONNUGHT LAB LTD.  
XX  
XX Mordin AD, Oomen RP, Wang J;  
PI  
XX  
XX WPI; 2000-350688/30.  
DR P-PSDB; AAY92818.  
XX  
XX Chlamydia antigens and the proteins they encode, useful for  
PT vaccinating against Chlamydia infections that affect the respiratory  
PT tract  
XX  
XX Claim 2; Fig 1; 226pp; English.  
XX  
XX The nucleic acids may be used for the recombinant production of the  
CC Chlamydia polypeptides (either in vivo or in vitro) according to standard  
CC recombinant DNA methodologies. The polypeptides may then be used to  
CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as  
CC C. pneumoniae, are pathogens responsible for upper respiratory tract  
CC infections such as community acquired pneumonia, acute respiratory  
CC disease and bronchitis and may be implicated in atherosclerotic changes  
CC and asthma. The nucleic acids may also be used as probes for detecting  
CC the presence of Chlamydia nucleic acids in samples (and therefore  
CC diagnose infections) and the proteins may be used as antigens for the  
CC production of antibodies that may be used to detect Chlamydia proteins  
CC in samples (e.g. via enzyme linked immunosorbant assay (ELISA)).  
XX

Sequence 2750 BP; 756 A; 532 C; 641 G; 821 T; 0 other;  
SQ  
Query Match 99.38; Score 2508.4; DB 21; Length 2750;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2515; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
OY 1 ATGAAGATTCGACCTCCGCTTTTATGATATCATAGTACCTAGCTTTATGTCGAAT 60  
|||||  
Db 101 atgaagatccaccctcccttctatgatacatcattagctaacgcttcttaagtcgaat 160  
OY 61 TTATTAGAGCTGCTACTACCGAAGACTATCGGCTAGCAATFACCTTCGATGAACTACA 120  
|||||  
Db 161 ttattagagctgctactaccggaagagttaacgcttagcaatagcttcgtaggaactaca 220  
OY 121 TCAACAAACAGCTTTTCTAGTAAACATCATCGGCTACAGATGGCCACCAATFATGTTTTT 180  
|||||  
Db 221 tcaacaacaagcttctctagtaaaacatcaccgctacagatgagcaccatattatgcttt 280  
OY 181 AAAGATCTGTAGTTATAGAAATGTAACCCAAACAGGGAACCTAGTCTACTAGTTGT 240  
|||||  
Db 281 aaagatctcgtagttatagaaatgtaacccaacagggaaactcagcttaactagttgt 340  
OY 241 TTTAAAAATGACGCTGACGCTGAGATCTAAATTTCTTAGAGGCGGATTTTCTTACACA 300  
|||||  
Db 341 tttaaaaatgacgctgacgctgagatcttaattcttagaggggatcttcttcaca 400  
OY 301 TTTAGCAATATCATGACACCGGCTTTCTGAGCTGCTATGGAAGTGAAGCAGCTAAT 360  
|||||  
Db 401 tttagcaatatacgtacacacacgctctctgagctgctatggaagtgagcagccta 460  
OY 361 AAGACAGTCAAGTATACAGATTTTGGCACTTTCTTCTTAATCCCGACAAGTACA 420  
|||||  
Db 461 aagacagtcacgtatatacagatcttcgcacttcttcttaacccacagcaagtaaca 520  
OY 421 GTGACTATGATTTGGAGCTATCATGTTAAAGGAATTTAAAGCTATTGGATATGAT 480  
|||||  
Db 521 gtgactatgatcttgaggactatcatatgtaaaaggaatttaagcctattgataatgat 580  
OY 481 AAGTATTTGATTCAGACATTTTCTCAACAGAGAGATGGCGGACCAATTAATTGTCAGGC 540  
|||||  
Db 581 aagtatctgattcagacaatttctcaacagagatgagcagaaactaattgtgcagcgc 640  
OY 541 TCCTTGAAGATCGCAACCAATTAAGTCCCTTTCTTATGGAATTAATGTTCTCAACAGCT 600  
|||||  
Db 641 tccttgaagatcgcaacaataaagcttcttcttcttcttgaatgattcttaacacgct 700  
OY 601 GCGGAGCGATTTACACAAAACCTCACATATCTCTGTGGGGAACCTATATTTCAG 660  
|||||  
Db 701 gcggagcgatctatcacaaaacctcaactatctctcgtggtgggaacctatcttcag 760  
OY 661 GGGAAATACAGCGCTACCGCTGCTGTGTAAGAGAGTCTATCGGATTTGCAAGATCTGGC 720  
|||||  
Db 761 gggaaatcacgacctacgctgctgtgtaagaggtgctatcgatctgagaaactcggc 820  
OY 721 ACCCTATCCATTTCTGAGACAGTGGGACATTTCTTGAAGCAATPAGATGAGAGCT 780  
|||||  
Db 821 accctatccattctcggagacagtcgcaactatcttgaagcaatcagataagagagct 880  
OY 781 ACAGAACCGTCTCTCATAGTGTATGATTTAGAACTAGCGCTAAGATPACTGCTTA 840  
|||||  
Db 881 acagaaacgctctctcatatgctatgatttaggaactagcgtaagataactcgctta 940  
OY 841 CGTCTCGCAGAGACATACGATTAATCTTTATGATCCGATTTACTGTATACAGATGACACA 900  
|||||  
Db 941 cgtctcgcagagacatacagatacttctatgatccgattactctgaacagatcgaca 1000  
OY 901 TCTGTGCTATATGTCATATTAATTAAGCCCTGATCTGGAGATACAAAGATFATAGC 960  
|||||  
Db 1001 tctgtgctatgctcctaataatagccctgatactcgtgagtaacaaagatacag 1060  
OY 961 GGAACCATAGTCTTTTCTGAGAGAACTCAGGAGCAACCTAAAGATGAGAAAGAC 1020  
|||||

Db 241 ttttaaatgacgcctgacgcctgagatctaaattcttagggagggtttcttcaca 300  
QY 301 TTTAGCAATATGATGACAAACGCGCTTGTGAGCTGCTATTGGAAGTGAACGCTAAT 360  
Db 301 ttttaacaatatgacgacaaacacgagcttcgagctgcatctggaagtgaacacgaat 360  
QY 361 AAGACAGTACGATTTCAGAGATTTTCGGCAGCTTTCTTTTAAATCCCGACAGATACA 420  
Db 361 aagacagtcacgtcttcaggatcttcgagacttcttcttaaatcccgacgaagtaca 420  
QY 421 GTGACTAATGATGAGGAGACTATCAATGTTAAAGCAATTTAAGCCTATTGATATGAT 480  
Db 421 gtgactaaatgattgggagctatacaatgattaaaggaatttaagcctattgataatgct 480  
QY 481 AAGGATTTGATTCAGAGCAATTTTCAACAGAGATGGCGGAGCAATTAATTGTGCAGGC 540  
Db 481 aaggtattgattcagagcaattctccaacagagatgagggagcaatttaattgtgcaggc 540  
QY 541 TCCCTGAAGATGCAAAACATTAAGTCCCTTTCTTTTATTTGAAATAGTCTTCAACAGCT 600  
Db 541 tccctgaagatgcgaacaataaagtccttcttcttattggaataatgcttcttcacaacgt 600  
QY 601 GCGGAGCGATTTCATACCAAAAACCTCACACTTCTGTGGGGAACCTGATTTTCAG 660  
Db 601 ggcggagcgattcataccaaaacctcaactatctctgtggtgggaaactctattcag 660  
QY 661 GGGATPACGCGCCACGCGCTGCTGTTAAAGAGTGCATATGCGATTCGATTCACACTGCGC 720  
Db 661 gggatpaccgcccacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 720  
QY 721 ACCCTATCCATTTCGAGACAGTGGCGACATTAATCTTTGAAAGCAATTCAGATAGAGCT 780  
Db 721 accctatccattctcgggagacggtgagacatatactcttgaaaggacaatacgttagagct 780  
QY 781 ACAGGACCGCTCTCATATGCTATTTGATTTAGAACTAAGCCTTAAGATTAACGCTTA 840  
Db 781 acaggaccgctctcatatgctatctttagaagaaacgcttaagaataactcgttaca 840  
QY 841 CGTGTGCGCAAGACATTCGATATCTTTTATGATCCGATCTAGTAAACAGAGTGCACA 900  
Db 841 cgtgtgcgcaagacatacagataacttctttagatccgataactcgttaacagagacga 900  
QY 901 TCTGTGCTGATGCTCTCAATATTAATAGCCCTGATCTGAGATTAACAAGATTAAGC 960  
Db 901 tctgtgctgactgctcctaataataatagccctgatactgagatacaagaagtatacgt 960  
QY 961 GGAACCATATGCTTTTCGAGAGAAAGTCAAGGAGGAGCAAGCTTAAGATTAAGAAGAAC 1020  
Db 961 ggaaccatatgcttcttcgagagaaagctcacgagagcgaaagctaaagatlgagaagac 1020  
QY 1021 CCGACTTTCATTAATTAATCAAAATGTGCTTTTAAATGGAGCTGTAGTTTAAAGGT 1080  
Db 1021 cgcacttctaataatcttcaaaaatgttgcctttaaataatggagactgttagttttaaaggt 1080  
QY 1081 GATGTGCTTTTAAAGTGCAGACGGTTTCTCTCAGATGCAAACTTAAGTTGATTAAGAT 1140  
Db 1081 gatgtgctttttaaagtgcagacggtttctctcagagatgcaaaacttaagttttagat 1140  
QY 1141 TTAGGAGAGTGTGTTGTTGCAAAACCGAAAGTATCGAGTTAACGAATTTGGAATTAAT 1200  
Db 1141 ttaggagagtgctgtgtgcaaaacccgaaagttacgagtttlaacgaatttggaaattaat 1200  
QY 1201 ATAGACTCTCTAGAGAAAGGAAAGATTAATACTCAGTGGCGCAAGCTCAGAAAGAT 1260  
Db 1201 atagactctctcagagaaagggaaagataaaactcagtgctgccaagctcagaagaagt 1260  
QY 1261 ATTGCTATAGATGTCCTGTTGTTGATGCGCAATTAAGCATGAGAGATTTTATCAAAATGCG 1320  
Db 1261 atctglatagatgctcctgctgtlactggaatcagatgagagtgcttctatcaaaaatgct 1320  
QY 1321 TTTTGAATGAGAGCAATTCCTATGATGGATTTCTTGAGTTGAGCTGGGGAANAACATC 1380  
Db 1321 ttttgaatgagagcaattcctatgagtgagattcttgagtttagatgctgggaaagacatc 1380

QY 1381 GTGATTTTCGAGATTTCTCGCAGTATTAATGCTGTACATCTCCGATNGCATACAGGA 1440  
Db 1381 gtgatttctcgagatttctcgagatTTAATGCTGTACATCTCCGATNGCATACAGGA 1440  
QY 1441 AAGTGACATCAATTTGGTCTACTAGTATAAAGAAAGTACGAGTTTCTTTGGGCAAGCA 1500  
Db 1441 aagtgacataatattgctactcagatgataagaaagcaagcttcttggcgaaagcg 1500  
QY 1501 AGTTTAAATCCACTGCTGAGCAGAGAGCTCCGTTAAGTTCTTAATCTTTGGGTTCT 1560  
Db 1501 agttttaaaccacgctgagcagagagctccgttagttccctaactcttcttggggttct 1560  
QY 1561 TTTATAGATGTTGTCCTTCCAAAATTTTATAGAGCTGATGACTGAAGAGTCTCCTTAC 1620  
Db 1561 tttatagatgcttcttcccaagaattcttaagagctagagctagaaaggtctctctac 1620  
QY 1621 GAAAAGATTTTGGGTTGACAGCATTTTCCAAATGTTTTCATAGAGAGCGTCTGAAAT 1680  
Db 1621 gaaaagagatttgggttgacagcatcttccaatgttttgcataggagcgctgtgaaat 1680  
QY 1681 CAAAGGAATTTCCGTCATGTTGAGTGAAGGTGCTGATAGTGTCTACGACGAGATCGC 1740  
Db 1681 caaaggaaattccgcatagtgagtgagtgctgtagtgctagcagcagagatgctcg 1740  
QY 1741 GGTGGATPACCTTGCTGCTGGGTTTGGCTCAGCTTTTGGCGTGACAAAGACTACTTT 1800  
Db 1741 ggtggatpacccttgctgctggggttggctcagcttttggcgtagacaaagactacttt 1800  
QY 1801 ATGAATPACCAATTTTGGCAAAAGACTACGAGATCTTTACGTTTTCAGCACATGCTTTC 1860  
Db 1801 atgaatpaccatcttcgaaagacctaagcagagatcttaacgtttgcagagatgcttcc 1860  
QY 1861 CTATCTCTGTGTGATATCTTTTATAGAGAGAGCACTCCGAGATCTCTGTTGCT 1920  
Db 1861 ctatctctgtgtgatatcttctttagagagagagctccgcgaaatctcgttgcct 1920  
QY 1921 TATGTTTCCAAAGCTCGCGCTGCTTCTTATGGGCACTAGCTACGCGCATACGAT 1980  
Db 1921 tatgttccaagaactcgcgctgctcttctatgycagtgatagtcagccatcaggt 1980  
QY 1981 CATGCCATGACCGAGTCTTACCCCCCCCCCGACGCTCTCGACGATCATACT 2040  
Db 1981 catgccatgacccgagctctctaccccccccccgacgctctcgcgagatcatact 2040  
QY 2041 TCTTGGGAGGATGCTGGGCTGGAGAGTGGGAACTCGATGCTGTAAATATAC 2100  
Db 2041 tcttgggagatagctgggctggagagtggaactcgatgctgttaaaatatac 2100  
QY 2101 AGCGGACAGATTTTTCGAGAGTACACTCCATTTGTAAAGTCCAGGCTTTTACTCG 2160  
Db 2101 agcggacagagatttccaagaagtaactcatttgaataaagctcaagctgtttaagct 2160  
QY 2161 CGCCAAGATAGCTTTGTTGAACTAGAGCTATCAGTCTGATTTTATGATTCGACTCT 2220  
Db 2161 cgccaagatagcttcttgaactagagctacagtcgattttagatctgcactct 2220  
QY 2221 TATACTTTGCAATCTCTTGTGATTCAGATTAAGAAAGGTTTTCAGAGCAATTAAT 2280  
Db 2221 tataactttgcactctctcttgaatcaagtttagaagaaagcttgcagagcaataat 2280  
QY 2281 CATGTTGATGATGATTTCTCAGATGTTTGTCTAGTAACCCCAATGTTGCACTAAC 2340  
Db 2281 catgtttagatgagatattcttcagaatgttttgtagtaacccaatgtacgctaac 2340  
QY 2341 CTACTTTCCAAACCAAGGAGTGAAGACCAAAAGTTCGAATTAAGACAGAGCTGCT 2400  
Db 2341 ctacttccaacaaaggagtgtaagaccaaaggttcgaacttagcaagaagctgct 2400  
QY 2401 ATTGTTAGGCTCAGGTTTGTGATCTTTGGAGTGCAGAGCTTTTGGGAACCTTT 2460  
Db 2401 attgttaggctcaggttcttcagatcttggagagctgcagagagcttcttgggaactt 2460

[illegible]

Query Match	Best Local Similarity	Matches 2515;	Conservative	Score 2508.4;	DB 21;	Length 2526;
XX	Chlamydia pneumoniae.					
OS	Chlamydia pneumoniae.					
XX	WO200024765-A2.					
PN	04-MAY-2000.					
XX	28-OCT-1999;	99WO-CA00992.				
XX	28-OCT-1998;	98US-0106034.				
XX	28-OCT-1998;	98US-0106039.				
PR	28-OCT-1998;	98US-0106042.				
PR	28-OCT-1998;	98US-0106044.				
PR	29-OCT-1998;	98US-0106072.				
PR	29-OCT-1998;	98US-0106073.				
PR	29-OCT-1998;	98US-0106074.				
PR	29-OCT-1998;	98US-0106087.				
PR	02-NOV-1998;	98US-0106587.				
PR	02-NOV-1998;	98US-0106588.				
PR	02-NOV-1998;	98US-0107034.				
PR	02-NOV-1998;	98US-0107035.				
XX	(CONN-) CONNAUGHT LAB LTD.					
XX	Murdin AD, Oomen RP, Wang J;					
PI	WPI: 2000-350688/30.					
XX	P-PSDB; AAY92818.					
DR	Chlamydia antigens and the proteins they encode, useful for					
XX	vaccinating against Chlamydia infections that affect the respiratory					
XX	tract					
PS	Claim 2; Fig 1; 226pp; English.					
XX	The nucleic acids may be used for the recombinant production of the					
XX	Chlamydia polypeptides (either in vivo or in vitro) according to standard					
CC	recombinant DNA methodologies. The polypeptides may then be used to					
CC	vaccinate against Chlamydia infections in mammals. Chlamydia, such as					
CC	C. pneumoniae, are pathogens responsible for upper respiratory tract					
CC	infections such as community acquired pneumonia, acute respiratory					
CC	disease and bronchitis and may be implicated in atherosclerotic changes					
CC	and asthma. The nucleic acids may also be used as probes for detecting					
CC	the presence of Chlamydia nucleic acids in samples (and therefore					
CC	diagnose infections) and the proteins may be used as antigens for the					
CC	production of antibodies that may be used to detect Chlamydia proteins					
CC	in samples (e.g. via enzyme linked immunosorbant assay (ELISA)).					
XX	Sequence 2526 BP; 693 A; 493 C; 601 G; 739 T; 0 other;					
SQ						
Query Match	Best Local Similarity	Matches 2515;	Conservative	Score 2508.4;	DB 21;	Length 2526;
99.3%;	99.6%;	0;	Mismatches 11;	Indels 0;	Gaps 0;	
1	ATGAAGATTCCACCTCCGCTTTTATTTATGATATATATAGTACCTACCGCTTCTATGTCGAAT	60				
1	atgaagattccactccgcttttattgatatcatcattagctaccgcttctctatgctgaat	60				
61	TTATTAGGAGCGTCTACTACCGAAGAGGTATGCGTACGATGATGCGTGAAGAACTACA	120				
61	ttattaggaagcgtcactaccgaaagataltcgtctagaacaagctcgaatggaactaca	120				
121	TCACACAAAGCTTTTCTAGTAAACATCATCGGTACAGATGGACACCAATATATGTTT	180				
121	tcaacaacaagctttctcagtaacaacatcatcgtctacagatgagaccacaattatgtttt	180				
161	AAAGATTCTGTAGTTATAGAAAATGTACCCAAACAGGGGAAATCTAGTCTACTAGTGT	240				
161	aaagttctgtatgtatagaanaatgtaccacaacaagggaaaccctcactactagttt	240				
241	TTTAAAAATGACGCTGACGCTGAGATCTAAATTTCTTATGACGAGGGGATTTTCTTTCACA	300				
241	tttataaaatgacgctgacgctgagatctaaatttcttattgacgaggggattttcttttcaca	300				

outer membrane proteins of *C. pneumoniae* or nucleic acids encoding these proteins

Claim 6; Page 49-50; 115pp; English.

This DNA sequence codes for the novel 89.7 kDa surface exposed protein Omp7 (see AAW8420) of the human respiratory pathogen *Chlamydia pneumoniae*. By generating antibodies against *C. pneumoniae* outer membrane complex, a polyclonal antibody (PAB 150) was obtained which reacted with outer membrane proteins. The antibody was used to identify the genes (see AAX0616-27) encoding Omp4-Omp15 proteins (see AAW8417-28) in an expression library of *C. pneumoniae* DNA. The genes are situated in 2 gene clusters: Omp12, 11, 10, 5, 4, 13 and 14 in one cluster and Omp6, 7, 8, 9 and 15 in the other, and encode polypeptides of about 89.6-100.3 kDa and about 56.1 kDa. The invention provides a new species specific test for identifying mammals (including humans) infected with *Chlamydia pneumoniae*. The test comprises detecting antibodies specific for Omp4-Omp15 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used in the diagnosis of *C. pneumoniae* infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with *C. pneumoniae*.

Sequence 2526 BP; 696 A; 495 C; 597 G; 738 T; 0 other;

Query Match 100.0%; Score 2526; DB 20; Length 2526;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAGAGATTCACCTCCGCTTTTATGATATAGTACACGCTTCTATGTCGAT 60
DB 1 atgaagattccaccctccgttcttatgatacatagtaactacgtcttctatgtgaat 60
QY 61 TTATTAGAGCTGCTACTACCGAAGAGCTATGCGCTAGCAATAGCTCGATGAATACA 120
DB 61 ttattagagctgctactaccgaagagctatcgctagaataagctctgagtaactaca 120
QY 121 TCAACAACAAGCTTTTCTAGTAAACATCATGCGCTACAGATGGACCAATTATGTTTT 180
DB 121 tcaacaacaagcttctctagtaaacatcatcgctacagatgycacaactatgctttt 180
QY 181 AAGAGTTCGTAGTATAGAAAATGTACCCAAACAGGGAACTGACTACTAGTTGT 240
DB 181 aagagttctgtatagtaaaaatgtacccaaacagggaaactgactactagttgt 240
QY 241 TTTTAAATGAGCTGCACGTGAGATCTAAATTTCTTGGAGGGGATTTCTTTTACA 300
DB 241 ttttaaatgagctgcacgtgagatcttaatttcttggaggggatttctttttaca 300
QY 301 TTTTACCAATATGATGCAACACAGGCTTCTGAGCTGCTATTGGAAGTGAACAGCTAAT 360
DB 301 ttttaccaatatgattgcaacacaggcttctgagctgctatttggaagtgaacagctaat 360
QY 361 AAGACAGTACGCTTATACGAGATTTTCGACATTTCTTTCTTAAATCCCACAGACTACA 420
DB 361 aagacagtcacgttatacagatcttcgcactcttcttctaatacccacaaatata 420
QY 421 GTGACTAATGATTTGGAGACTATCAATGTTAAAGGAATTTAAGCTATTGATATATGAT 480
DB 421 gtgactaatgatgttgagactatcaatgtttaaggaatttaagctattgataatgat 480
QY 481 AAGGATTTGATTCAGACAAATTTCTCAACAGAGATGGCGAACAATTATTTGCGAGCG 540
DB 481 aaggatttgcattcagacaattctcaacagagatgycgagcaatattgtgcagcg 540
QY 541 TCCCTGAAGATGCGAAACAATAAGTCCCTTTCTTTATTTGGAATAGTCTTCAACAGCT 600
DB 541 tccctgaagatcgcaaaataaagtccttcttcttattggaataagtgcttccaacagt 600
```

```
QY 601 GCGGAGCATTCATACCAAAAACTCACACTATCTTCTGTTGGGAACTCTATTTTCAG 660
DB 601 gcgagagcatctcataccaaaaactcacactatctctcgtggyggaactctatttcag 660
QY 661 GGGAAATACAGCCCTACGGCTCTGTTAAAGAGTGTATCCGATTCGACACTCTGCG 720
DB 661 gggaaatacagccctacggctctgtttaaagagtgtatccgatctcgaattgcagactcgcg 720
QY 721 ACCCTATCCATTTTGGAGACAGTGGCGACATTTATCTTTGAAGGCAATACGATAGAGCT 780
DB 721 accctatccattcttgagacagctgycacatattcttgaaggaatacagataagagct 780
QY 781 ACAGGACCGCTCCATATGTGCTATGATTTAGAGATACGCGCTAATTAATCTGGTTA 840
DB 781 acaggaccgctccatattgtctatgtattagaacacagcctaagataactgcgtta 840
QY 841 CGTGTGCGCAAGACATACGATATACCTTTTATGATCCGATTAAGTAAACAGATGACA 900
DB 841 cgtgtcgcgaagacatacgatatactattatgatacgcattatcgatattcgttaacagatcgaca 900
QY 901 TCTGTTGCTGATGCTCTCAATATTAATTAACCTGATACGAGATATACAAAGATATAG 960
DB 901 tctgttgctgatgctctcaatatttaatttaacctgatacgagatatacaaaagatatacg 960
QY 961 GGAACCATAGTCTTTCTGAGAGAAAGCTCACGAGGCGAAGCTTAAAGATGAGAAAC 1020
DB 961 ggaaccatagcttttctgagagaaagctcacgagggcgagagcgaagataagagaagaaac 1020
QY 1021 CGCACTTTCTAAATTAATCTCAAAAATGCTTTTAAATATGCACTGATTTTAAAGAT 1080
DB 1021 cgcactttctaaatttaattcttaaaatgcttctttaaataagggactgattttaaagatt 1080
QY 1081 GATGTCGTTTAAAGTACGAAAGCTTCTCTAGATGCAACCTTAAGTTATGAT 1140
DB 1081 gatgtcgttttaagtacgaaagcttctctagatgcaaccttaagtattatgatt 1140
QY 1141 TTAGGAGACGTGTTGGTTGCTCAAAACACCGAAATATGAGTTAACAATTTGAATTAAT 1200
DB 1141 ttaggagacgtgttggttgctcaaaaacccgaaatattgagtttaacaatttgaaatatt 1200
QY 1201 ATAGACTCTCTCAGACAGCGGAAAGATTAACATAGTGTGCTCCACACTCAGAAAGAT 1260
DB 1201 atagactctctcagacagcggaagataaactaagtgctgcacactcagaagaat 1260
QY 1261 ATTGCTATGATCGTCTGTGACTGCAATTAAGGATGAGAGTTTATCAAAAATGAGC 1320
DB 1261 attgctatgatcgctgtgactgcaatttaaggatgagagtttattcaaaaatgagc 1320
QY 1321 TTTTGAATGAGAGACCATTCATATGATGGATTTCTTGAATGATCTGGAAAGACATC 1380
DB 1321 ttttgaatgagagaccattcatatgattggatttcttgaatgattctggaaagacatc 1380
QY 1381 GTGATTTCTGACAGATTCGACAGTATTAATGCTATCAATTCOCGTATGCTATCAGGGA 1440
DB 1381 gtgatttctgacagatctcgacagatataatgctataactccgtatgctatacaggga 1440
QY 1441 AAGTGACAAATCAATTTGGTCTACTGTATGAAGAAGACATACGTTTCTTGGGCAAGCAA 1500
DB 1441 aagtgacaatcaatttggctactgtatgaagaagacatcacggttcttgggcaagcaa 1500
QY 1501 AGTTTAAATCCACCTGCTAGACAGAGGCTCGTTAGTTCCTAATCTTCTTGGGGTTCT 1560
DB 1501 agttttaaattccacctgctagcagagggctcgttagcttccctaactcttcttggggttct 1560
QY 1561 TTTTATGATGTTGCTGCTTCCAAAAATTTTATAGAGCTAGTACGTAAGAGTCTCTTAC 1620
DB 1561 ttttatagatgttctgcttccaaaaatttataagagctagactagaagttctctcttc 1620
QY 1621 GAAAGAGATTTTGGGTGCGAGGATTTCCATGTTTTCATAGAGGCGGTGCTGTAAT 1680
DB 1621 gaagagatcttgggtgctgagggatcttccaatgttctcacaagagcggtcgtgaatat 1680
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 05:06:28 ; Search time 800.41 Seconds  
(without alignments)  
1981.584 million cell updates/sec

Title: US-09-446-677B-7  
Perfect score: 2526  
Sequence: 1 ATGAGCAATTCACATCCGCTT.....GTACCAAAATCAATTTAG 2526

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: N\_Geneseq\_0601.\*  
2: /SIDSL/gcgdata/geneseq/NA1980.DAT.\*  
3: /SIDSL/gcgdata/geneseq/NA1981.DAT.\*  
4: /SIDSL/gcgdata/geneseq/NA1982.DAT.\*  
5: /SIDSL/gcgdata/geneseq/NA1983.DAT.\*  
6: /SIDSL/gcgdata/geneseq/NA1984.DAT.\*  
7: /SIDSL/gcgdata/geneseq/NA1985.DAT.\*  
8: /SIDSL/gcgdata/geneseq/NA1986.DAT.\*  
9: /SIDSL/gcgdata/geneseq/NA1987.DAT.\*  
10: /SIDSL/gcgdata/geneseq/NA1988.DAT.\*  
11: /SIDSL/gcgdata/geneseq/NA1989.DAT.\*  
12: /SIDSL/gcgdata/geneseq/NA1990.DAT.\*  
13: /SIDSL/gcgdata/geneseq/NA1991.DAT.\*  
14: /SIDSL/gcgdata/geneseq/NA1992.DAT.\*  
15: /SIDSL/gcgdata/geneseq/NA1993.DAT.\*  
16: /SIDSL/gcgdata/geneseq/NA1994.DAT.\*  
17: /SIDSL/gcgdata/geneseq/NA1995.DAT.\*  
18: /SIDSL/gcgdata/geneseq/NA1996.DAT.\*  
19: /SIDSL/gcgdata/geneseq/NA1997.DAT.\*  
20: /SIDSL/gcgdata/geneseq/NA1998.DAT.\*  
21: /SIDSL/gcgdata/geneseq/NA1999.DAT.\*  
22: /SIDSL/gcgdata/geneseq/NA2000.DAT.\*  
23: /SIDSL/gcgdata/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2526	100.0	2526	20	AA06819
2	2508.4	99.3	2526	21	AA28691
3	2508.4	99.3	2750	21	AA28690
4	2485.4	98.4	1230025	20	AA09190
5	1425.2	56.4	1830	20	AA06830
6	341	13.5	2950	21	AA30847
7	341	13.5	3200	20	AA06816
8	340.2	13.5	273254	21	AA06814
9	338.6	13.4	3000	21	AA27021
10	326.6	12.9	2787	20	AA06820
11			2784	21	AA30848

12	313.6	12.4	273254	21	AA081914	Chlamydia pneumoniae
13	309.4	12.2	2815	20	AA06817	Chlamydia pneumoniae
14	297	11.8	2957	20	AA261509	DNA encoding the C
15	295.4	11.7	2757	20	AA06821	Chlamydia pneumoniae
16	295	11.7	2950	21	AA06849	Chlamydia antigen
17	287.4	11.4	3000	20	AA06828	Chlamydia pneumoniae
18	280	11.1	2793	20	AA06823	Chlamydia pneumoniae
19	278.4	11.0	3000	21	AA06853	Chlamydia antigen
20	275.4	10.9	3050	21	AA06838	Chlamydia pneumoniae
21	269.6	10.7	2787	20	AA06822	Chlamydia pneumoniae
22	268	10.6	2950	21	AA06851	Chlamydia antigen
23	266.6	10.6	2784	21	AA06852	Chlamydia antigen
24	256.8	10.2	2808	21	AA06839	Chlamydia pneumoniae
25	252.6	10.0	2790	21	AA06854	Chlamydia antigen
26	250.8	9.9	2784	21	AA06850	Chlamydia antigen
27	192.8	7.6	3050	21	AA261508	DNA encoding the C
28	189.6	7.5	2838	20	AA06827	Chlamydia pneumoniae
29	180.6	7.1	3150	21	AA06832	Chlamydia pneumoniae
30	153.6	6.1	2950	21	AA06835	DNA encoding Chlam
31	150.4	6.0	3052	20	AA06818	Chlamydia pneumoniae
32	130.4	5.2	840	20	AA06824	Chlamydia pneumoniae
33	102.8	4.1	936	22	AA06825	Oligonucleotide D1
34	102.8	4.1	936	22	AA06825	Oligonucleotide D1
35	102.8	4.1	936	22	AA06825	Oligonucleotide D1
36	102.8	4.1	936	22	AA06825	Oligonucleotide D1
37	102.8	4.1	936	22	AA06825	Oligonucleotide D1
38	102.8	4.1	936	22	AA06825	Oligonucleotide D1
39	100.8	4.0	936	22	AA06825	Oligonucleotide D1
40	100.8	4.0	936	22	AA06825	Oligonucleotide D1
41	100.8	4.0	936	22	AA06825	Oligonucleotide D1
42	100.8	4.0	936	22	AA06825	Oligonucleotide D1
43	100.8	4.0	936	22	AA06825	Oligonucleotide D1
44	100.8	4.0	936	22	AA06825	Oligonucleotide D1
45	99.6	3.9	1200	20	AA06829	Chlamydia pneumoniae

## ALIGNMENTS

RESULT 1	AA06819	standard; DNA: 2526 BP.
ID	AA06819	
XX	AA06819	
AC	26-APR-1999	(first entry)
XX		
DE	Chlamydia pneumoniae surface exposed protein Omp7 DNA.	
XX		
KW	Omp7: outer membrane protein 7; surface exposed protein; antigen; infection; diagnosis; vaccine; atherosclerosis; asthma; ss.	
XX		
OS	Chlamydia pneumoniae.	
XX		
PN	MO9858953-A2.	
XX		
PD	30-DEC-1998.	
XX		
PF	19-JUN-1998; 98WO-DK00266.	
XX		
PR	23-JUN-1997; 97DK-0000744.	
PA	(BIRK/) BIRKJUND S.	
XX	(CHR/) CHRISTIANSEN G.	
PI	Birkjund S, Christiansen G, Knudsen K, Madsen A;	
PI	Mygind P;	
XX		
DR	WPI, 1999-105610/09.	
XX	P-PSDB; AAW88420.	
PT	Species-specific test for identifying mammals infected with Chlamydia pneumoniae - comprises detecting antibodies specific for	

**THIS PAGE BLANK (USPTO)**









Db 2430 CTTGAACCAAGATATACATATCTACTGTTAAAGAGAGCTGGGGGAATGATGTTTC 2489  
QY 2047 GGAGGATATGCTGGGGCTGGAGAGCTGGAGACTCCAGTGTCTGTTGAAATACAGCGCC 2106  
Db 2490 GCTTTAGAAATGCGTGGAGAGCTCCGATTTGCTTAGAT-----GAA 2531  
QY 2107 AGAGATTTTTCGAGAGACATCCATTTGTAAGGTCGAAGCTGTTCAGTCCGCCAA 2166  
Db 2532 AGTGCCTCTTTTGGAGAGATGATGCCCTTCATGAATTTGACGTTGTCTATGACATCAG 2591  
QY 2167 GATAGCTTTTGAACATAGAGACTATCAGTCGTATTTAGTATTCGATCTTTATAC 2226  
Db 2592 GAAGGTTTAAAGACAGAGAGACAGAGCTCGTAATTTGGAAGTACCGCTCTTGATGAT 2651  
QY 2227 CTTGGCATTCCTTGGATTCAGATTTAGAGAAAGGTTTTCAGAGCAATAT-----TAT 2280  
Db 2652 CTGGCTTACCTATGCGGATCCGATTTGATTAAGAAATCAGACTGCCAATGCAACGATC 2711  
QY 2281 CATGTTGTAGCATGATATTCCTCAGATGTTTGTGCTAGTAAACCCCAATGATCAGACTAC 2340  
Db 2712 AATCTACCTCTTGTATATCTGATGATCTGTTGCTGATAGTAAACCCGACTGTAGACATA 2771  
QY 2341 CTACTTTCCACCAAGAGGATTTGGAAGACCAAGGTTGCACTTACCAAGACAGCTGT 2400  
Db 2772 CTGCGAATTTAGGCGTGTATCTTGAAGAACTTCGATACGAATTTGGCAAGACAGCTTTA 2831  
QY 2401 ATTTGTCAGGCTCAGGTTTGGATCTTTGGAGCTGACAGAGCTTTTCCGAACTT 2460  
Db 2832 GTCTTCTGTCAGAGGACCATTTTGTCTTAACTCAATTTTGAAGCCCTTATGCCAATTT 2891  
QY 2461 GCGTTTGAATGCGGAGCTTCTCTAGTATTAATGATGATGCGGATAGCAAAATCAAA 2520  
Db 2892 TCTTTTGAATGCGTGTGATCTGCAATCTGCAATTAACATGATGACTTACAGCAAAATCA 2951  
QY 2521 TTTTGA 2525  
Db 2952 TTTCTA 2956

RESULT 14  
LOCUS A81835 2787 bp DNA 21-JAN-2000  
DEFINITION Sequence 9 from Patent WO9858953.  
ACCESSION A81835  
VERSION A81835.1 GI:6731868  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 2787)  
AUTHORS Maden, A. and Birkelund, S.  
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE  
JOURNAL Patent: WO 9858953-A 9 30-DEC-1998;  
MAUSEN ANNA SOFIE (DK); BIRKELUND SVEND (DK)  
FEATURES  
source 1. 2787  
Location/Qualifiers  
BASE COUNT 811 a 583 c 598 g 795 t  
ORIGIN

Query Match 13.4% Score 338.6; DB 9: Length 2787;  
Best Local Similarity 51.1% Pred. No. 2.4e-71;  
Matches 1014; Conservative 0; Mismatches 914; Indels 57; Gaps 7;

QY 559 AATAGTCCCTTCTTTTATTTGAAATAGTCTTCAACAGCTGGCGGAGCATTCATACC 618  
Db 841 AATCAGATGTACTCTTCAACAATATCATGACAAACGCGGAGAGACTATCTATG 900  
QY 619 AAAAAGCTACATATCTCTGCTGGGAAACTCTATTTTCAGGGGAAATACAGCGCTAC 678  
Db 901 AAAAAGCTGAACTGCTTCGAGAGACTTACCTATTTAGTAAATAGTGTCAATGGA 960

QY 679 G---CTGCTGTAAGAGAGGTGCTATCGGATTTGAGACTGTGGCCCTATCCATTCT 735  
Db 961 GGTACAGCTCTCAAGGTGGAGCCATGATATCGAAGATAGTGGGAATTTGATTTATCC 1020  
QY 736 GGAGACAGTGGGACATTTCTTTTGAAGCAATAGCATAGAGCTACAGAAACCGCTCT 795  
Db 1021 GCCGATAGTGGACATTTCTTTTATGAGGAAATACAGTCACTTCTACTCTCTGGAG 1080  
QY 796 CAT--AGTCTATTTGATTTTGAAGACTACGCTTAAGATTAATGCTGTACGTGCGCAA 852  
Db 1081 AATAGAAAGTATGATGCACTTACGAGAGAGTCAAGAGTACAGCTTTGCTGCTGCT 1140  
QY 853 GGACATATGATTTATGATTCGATTTACGTTTACAGAGATGATCAGTGTGCTGAT 912  
Db 1141 GGTAGAGCATCTACTTCTATGATTCCTATACAGATCTTCCACACAGTTACAGAT 1200  
QY 913 GCTCTCAATATTTAATAGCCCTGATACAGATTAACAAAGATATACGGAAACCATAGTC 972  
Db 1201 GCTTAAAGTTAATAGAGACTCCGAGATTTCTGCACTCAATATATACAGGAACATCATC 1260  
QY 973 TTTTCTGGAGAGAGCTCACGAGGAGAGAGTAAAGATGAGAAAGCCGACTTTAA 1032  
Db 1261 TTCCAGAGGAGAAAGTTATCAGAGACAGAGCGCGAGATTTCTAATAATCTTACCTGGA 1320  
QY 1033 TTACTTCAAAATGTTGCTTTTAAATGGAGCTGTAGTTTAAAGAGTATGTCGTTTA 1092  
Db 1321 CTACTACAGCTCTGATCTTTTCAGAGAGTACTCTTCTTTAAACATGAGAGTCTG 1380  
QY 1093 AGTGGCAAGGTTTCTCTAGAGATGCAACTGTAAGTATTAATGATTTAGGAGCTGT 1152  
Db 1381 CAGACTCAGAGCATTTCTCAACAGGAGATTTCTGCTGAATATGAGAGTACATAC 1439  
QY 1153 TTGTTGCAAAACCGAAGATTCAGATTAACGAATTTGGAATTAATATAGCTCTC 1212  
Db 1440 --TCTAAGAACTGCTGATCTAGCACCATTAACAAATTTGCTATTAATCAAGTCTTAA 1497  
QY 1213 AGAAGCGGAAAGATTAATCACTCAGTGTGCACTCAGAAAGATTTGATAGAT 1272  
Db 1498 GACGTTGCAAAAGAGCAAAATATAGAAACCAAGCTACGTCAAAATATGACTTTATCT 1557  
QY 1273 CGTCTGTTGTACTGCAATTTAGCATGAGAGTCTTTTATCAAAATGCTTTTGAATGAG 1332  
Db 1558 GGAACCATCATTTTATTTGACCCGAGCGGACAGTCTTATGAATAATCATATTAAGAAAT 1617  
QY 1333 GACATTTCTATGATGAGATTTCTTGATTAATGCTGGGAAGAGACATCTGATTTCTGCA 1392  
Db 1618 CCTCAGTCTTACGACATCTTAAAGCTCAAAAGCTCTGGAAGTGAACAAGCACCGCAGTG 1677  
QY 1393 GATTTCCGAGTATTAATCTGTACATCTCCGTATGAGTATCAGGAGAAAGTG----- 1446  
Db 1678 ACTCCAGATCTTATTAATGAGTGAAGAAATTCATTAAGCTATACAGGAACCTTGGGCCCA 1737  
QY 1447 ACAATCAATTTGCTATCATGATGAAGAAAGCTACGGTTTCTTGGCAAGCAAAAGTTT 1506  
Db 1738 ATTTGTTGGGGGACAGGGGCTTCTACGACTGCAACCTTCAACGCTTAAACTGTGCTAT 1797  
QY 1507 AATCCACATGCTGAGAGAGAGGCTCGTATGTTCCAACTCTCTTTGGGGTCTTTTATA 1566  
Db 1798 ATTTCTAATCCGAGAGATGCGCTCTTTATGTCCTTAATAGCTTTATGATCATTTATA 1857  
QY 1567 GATGTTGCTCCCTTCAAAATTTTATAGAGCTAGTACGAGAGTGCCTTACGAAAG 1626  
Db 1858 GATATTAGCTCTCATATTTATTTAGAGACTGCAAGAGAGGTTGAGAGAGACCT 1917  
QY 1627 AGATTTGGGTTGACAGCATTTTCAATGTTTTCATAGAGAGCGTCTGTAATCAAG 1686  
Db 1918 GCTTTTGTGTGCTGAGTATCTTCAATTAAGATGATACAAAGATACAGACAGCC 1977  
QY 1687 AATTTCCGATGATGAGTGAAGAGGCTGTGATAGTATGAGTATGACAGAGATGCGGGTGT 1746  
Db 1978 GGGTTTCCGATTTTATGATGCGGTTATGTCATAGGAGAAACCTACATATCTTGTTCAGAT 2037



Db 210230 CAGCCATTCAGACACCGCATATGAGAAAGCAGCTATACCTCATTCGCCAATC----- 210280  
QY 2028 GAGGATTCATCTTCTGGGAGGATATGTCCTGGGAGGAGCTGGAGCTGAGCTGGC 2087  
Db 210281 -----CGAAGGTTCTTGGAGCAGACGATGATAGCTGGTGGATGGCTAGACCTTCC 210334  
QY 2088 TGTGTAATATACACGCGGAGAGGATTTTTCGAGAGTACATCTCATTTGTAAAGTCCA 2147  
Db 210335 TTTTGTTCTTCCAAACCAACATCCTTTTCAAGACCTTCATTCACAGATGAAAGTCGA 210394  
QY 2148 AGTGTCTTCTGCGGCGGAGATAGCTTTGTGAACACTAGAGCTATAGTCGATTTTAC 2207  
Db 210395 AATGTTTATGATCAACAAATAGCTTCTTCGAAACCTAGAGATGAGCCGTGGTTTAC 210454  
QY 2208 TGATTCGACATCTTATACCTTGGCATTCCTCTGTGATATCAAGTTAGAGAAAGGTTTGC 2267  
Db 210455 TATTTGAGAGGCTGCTTACCTTCGATCTCTGTTGGTGGAAATTTGTGACGGGGATAT 210514  
QY 2268 AGAGCAATA-----TTATCATGTTGTAGAGATATTCCTCCAGATGTTTGTCTAGTAA 2321  
Db 210515 CGGAGATTCCTACACCTATGATGCTCAGAGATTCCTTGTTCGATGCTATGCTAACA 210574  
QY 2322 CCCCAATGACGACTACCTTCTTCCAAACAGGAGCTGGAGAGCAAAAGTTTCAA 2381  
Db 210575 TCCCAATCTACAGCAGCTCTGTGATGAGCCGAGACTCTTGGAAATTCGCGGTGGCAA 210634  
QY 2382 CTTAGAGAGAGGCTGATGATGTTGTCAGGCTCAGCTTTTCATCTTTGGAGCTGCAGC 2441  
Db 210635 TCTTTCAAGACAGGCAATTTTCTAGAGGGGTAGCAACAACCTACGTTACAACTCAATTG 210694  
QY 2442 AGAGCTTTTGGGAACTTTGGCTTTGATGGGCGGGGATCTCTGATGATATAGTAA 2501  
Db 210695 TGAGCTCTTCGCGCATTTAGCGATGAGACTCGGTGATCTTCAAGAGACTACAAATGTA 210754  
QY 2502 TCGGCTAGCAAAATCAATTTTAC 2526  
Db 210755 TGTGGTACCACACTCGCATTTAG 210779

RESULT 13  
AE001587  
LOCUS AE001587 16448 bp DNA BCT 01-DEC-2000  
DEFINITION Chlamydia pneumoniae section 3 of 103 of the complete genome.  
ACCESSION AE001587 AE001363  
VERSION AE001587.1 GI:4376271  
KEYWORDS  
SOURCE  
ORGANISM Chlamydia pneumoniae CWL029.  
Chlamydia pneumoniae  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
REFERENCE  
AUTHORS 1 (bases 1 to 16448)  
Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W.,  
Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.  
Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
Nat. Genet. 21 (4), 385-389 (1999)  
JOURNAL  
MEDLINE 99206606  
PUBMED 10192388  
TITLE 2 (bases 1 to 16448)  
Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,  
Grimwood, J., Davis, R. W. and Stephens, R. S.  
Direct Submission  
Submitted (01-DEC-1998) Program in Infectious Diseases, University  
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
FEATURES  
source  
1. 16448  
/organism="Chlamydia pneumoniae CWL029"  
/strain="CWL029"  
/db\_xref="taxon:115713"  
misc\_feature  
172..2957  
/note="Polymorphic Outer Membrane Protein G Family;  
contains frame shift"  
gene  
172..705  
/gene="pmp-3.1"  
/note="Cpn0014"  
/db\_xref="GI:4376274"

gene  
615..2957  
/gene="pmp-3.2"  
/note="Cpn0015"  
3196..4971  
/gene="pmp-4.1"  
/note="Cpn0016"  
3196..5953  
/note="Polymorphic Outer Membrane Protein G Family;  
contains frame shift"  
4874..5953  
/gene="pmp-4.2"  
/note="Cpn0017"  
6302..7786  
/gene="pmp-5.1"  
/note="Cpn0018"  
6302..9139  
/note="Polymorphic Outer Membrane Protein G Family;  
contains frame shift"  
7787..9139  
/gene="pmp-5.2"  
/note="Cpn0019"  
complement(9380..11467)  
/gene="Cpn0020"  
complement(9380..11467)  
/gene="Cpn0020"  
/codon\_start=1  
/transl\_table=11  
/product="Predicted OMP [leader peptide: outer membrane]"  
/protein\_id="AAD18173.1"  
/db\_xref="GI:4376272"  
/translation="MKRCFLFASFLVGLNGSADALTHQEAVERKNSYLSHFESVGIY  
TLEDVLNHNHLRIQANKVYENTVGSLKIVAGNVAIVNRAKTLVQDYLEYED  
DSCLTNGFAPYPMFLGSMVTLPETIVIKGYISSEGRKCLDLSGVLEYSSD  
SLISCTKTLRYVCRIPILETPESIMPEIPKPPINPFGCGCGPGLSYLGGYPSIS  
KHFSSTPILDSFRKHGVMGPNVLRCSQKQVPEPNVNMKSYAHRALIDMAENHRYR  
HGDFCFTRHVAHVSSEYHLSDSWETVADIPNNEMLNKNGTPRVDCIWMNDYFEGYLT  
SSVKNVSQMANQELPYLRLROYPISYNTGVLYENIYECGLNFAFSDDHIVGENESS  
LRLARPKLHKTVPPLIGLISSTLSSLIYSDVPEISRHQSLSAKLQDLYRFLHK  
SYIQRHIIIEPVEFTTERPLAKNEDYIFSIODAFSLNLKAGIDPSYVSKTNP  
PPIHAKLMTHTHILSNTESKPPFPKTLACELSLPCKKNTVSLDAEMIKKHCWDMNT  
RHEHIGNDNVAMTLESLSHRKSYSLKCDRENITLDVSRPIDQLDLSRDNRLILK  
LFRPHPCWNTRLSLRYGWHRODTPNYLETOMILCTKIFEHWQLYGVYERRADSRFF  
FFLKLDKRRKPPF"  
complement(11484..13190)  
/gene="Cpn0021"  
complement(11484..13190)  
/gene="Cpn0021"  
/codon\_start=1  
/transl\_table=11  
/product="Predicted OMP [leader peptide]"  
/protein\_id="AAD18174.1"  
/db\_xref="GI:4376273"  
/translation="MGLFHLTFGLLCSPLISVAKRPESVGHILVITSTOQAL  
ATYELADAYGDHDFVLAKIGEDYLKOSIHSSDPQTRSTIIGAGLSSPADVLS  
QAMETADPQQQLVIAVSAVGHGKTSDDLFPALASPPYIKRLAAYRIANLKNKYI  
DHLSHFIRKLPELIIQLSAIFRLRLETESDAYVIDLLAAKSAIRSAFTALIGSYQO  
KREPLRLNLRLTASPODEATILVAGLKIDQOSYNNIKQKQKPDVNTYLAALAAI  
ALGREDALPVIKQALIERPRLVAILRHIPSEITIPALPIELKTKNSEAKLANYALA  
LLELSCDPRKLEIYTERLVQPHYMETALSSKSKRTIQNKRNVIIYPODQEBERL  
LSTTGLBEQIITFLRLPKKAYLPCTIKLLASQTYQALATTAISFLSHSTHOBADLL  
FOAKLPGEPRIIRAYADLAIVNLTPDKKRSIHDYAKLLIOETLLFVDTENRHPSP  
MPYLRYQVTPESRKLMDIETTLATSKSSSEDIRLLIOLMTEGDAKNFVLAGLLIKI  
VE"  
complement(13172..13762)  
/gene="maf"  
/note="Cpn0022"  
complement(13172..13762)  
/gene="maf"  
/codon\_start=1  
/transl\_table=11  
/product="Maf-type protein"  
/protein\_id="AAD18175.1"  
/db\_xref="GI:4376274"

```

/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="BA098485.1"
/db_xref="GI:8978649"
/translation="MDPERKNYDASAITVLEGLOAVBERPMIGDTGTLGHLHYE
VWNSIDEAMAGYCSRIDVRLIEDGQIVYDNGSIGIIEVHERSAGQREVSLAEV
LTVLHAGKFDKDSKVLHGVGCVNLSLKLALYAFVDRKCYOMERSRGIPAT
PLOWYSVDRQGTIVFPDPIRSTCTPDSILMKRLBELAFNRGITYFEEDRQV
SPKVFPEVFGIOSFVSYLNONKESLFSPRYICGTRVGDDEFEFEALOMNSGY
EIVSYANNPTPRQGTHTLGFSPALTRVINTYIKRANKLNKLALITGDRGLTA
VISVNPPEQGTOKLNSDVSAVOOVGHALTFEENQIARMTVDKFLVA
QAREAAKARELILRKSALDSARLPKILIDLEDPKCEMYIVEGDSAGSARQGD
RRFOAIIPIRGIKILNEKARLOKIFQNDIEITIAAGCGIGADNFMSKLRRIII
MTDADVDSHRLITLTFEYRHMALLENCEVYIAOPILVSKKDEKRYILSEKMD
SYLMIGTNESSILFKSTERREGALESPFINTLADVESPINTLEKKAIPSELEY
KEGIGYPLYIAPATMGQGRYIKXSDEKEBALAOEETHKRIIELYKRVAFYDIOQ
LKEYGDISSILPQKNEIVIGNEDSPSCSTCLEEVNLYNLKRLKGIETIRYG
LEGEMADQMDTMMNPORLIIHVSILDAVEADHIFMLMGEVPPRRFEISHALSI
RINLNDI"
complement(10812..11159)
/gene="CpJ0276"
/complement(10812..11159)
/gene="CpJ0276"
/codon_start=1
/transl_table=11
/product="Cti91 hypothetical protein"
/protein_id="BA098486.1"
/db_xref="GI:8978650"
/translation="MFLRRKRKGGSOVONKRTASPIKHAHYHLYLOELOKIMARP
HDAIDANVOYFRDKRKMSSQAIIGFRDHIILVKVYNSLYALLKOTPONDILMSLYOVA
SHVOIREIOFLG"

```

Query Match 13.5% Score 341; DB 2; Length 300650;  
 Best Local Similarity 50.1%; Pred. No. 7.6e-72;  
 Matches 1095; Conservative 0; Mismatches 1015; Indels 75; Gaps 7;

```

QY 396 TTTTCTTAATCCCGAGCAAGTACGACTAATGATGGAGCTATCAATGTTAAAG 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208616 TTAATGATGATCCATACATACATACGCTACGACAGAGCGATACGATGAAAG 208675

QY 456 GAATTTAAGCTAATGATAGCTATGATTCAGACCAATTTCTCAACAGAGA 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208676 CACGTCGACTACGAAACAATTTCTATATTGAAAGGAATGACGCAAAACTAC 208735

QY 516 TGGGGGAGCAAT-----AATTGACAGCTCCCTTGAAGATCGCAAAATA 560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208736 TGGCGGTGCGATGCAACACCAAGCGAGTGATCTCTCAAGCATATCTCTAACAA 208795

QY 561 TAACTCCCTTTCTTTATGGAAGTATGTTCTCAACAGCTGGCGAGCGATTACCA 620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208796 TAAACTCTGATCTTGTCTCAAGCTAGCAAGAAACAGCGGTGGCGCATCCATGCTAA 208855

QY 621 AAACCTCACATATCTCTGTTGGGAAACTCTATTTCAGGGGAATACAGCGCTACGGC 680
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208856 AAACCTAGCCCTTCTCTGAGGCTTTACAGAGTTTCTAGAAATATATGTCATCAGC 208915

QY 681 TCGTGGAAGAGAGTGTATCGGATTCAGACTGCGACCGCTATCCATTTTGGAGA 740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208916 AACTCTTAAGGGGGGTCTATCAGATCGATGCTCAGAGAGCTCACTCTTTCGAGA 208975

QY 741 CAGTGGGACATTAATCTTGAAGCAATACGATAGAGCTACAGAAACCGTCTCT----- 795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208976 GACAGGAACATTAATCTTGAAGAAATACCCCTACAAACAGCGAATACCGATACCTCC 209035

QY 796 -----CATAGTCTATTGATTTAGAACTAGCGCTAAGATACTCGCTTACGTCTGCGCA 851
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209036 TAAACGTAAATGCGATCAACATAGGAAGTAACGGAATTCACGGAATATACGGGTCTGTA 209095

QY 852 AGGACATACGATATACCTTTTATGATCCGATTAAGTAACAGAGATCGACATGTGTGCTGA 911
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209096 AATATCATCAATTTTCTTATGATCCATCACTTCAGAAAGAACTCAT-----CAGA 209149

QY 912 TGCTCTCAATTAATAGCCCTGATGAGATTAACAAAGATATAGGAGACCATAGT 971

```

```

Db 209150 CGTATTAAGATTAATTAACCGGCTGCGGAGCTCTCAATCATATCAAGAAACGATCT 209209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 972 CTTTCTGAGAGAAAGCTCAGGAGCGAGAAAGCTAAATATGAAAGAACCCGACTTCTAA 1031
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209210 ATTTCTGAGAAACCCCTAACAGCAGATGAACCTTAAGTCTGTCATTTAAAAATCTTC 209269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1032 ATTACTCAAAATGTGCTTTTAAAAATGGGACCTGACTTTTAAAAAGGTGATGTCGTTT 1091
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209270 ATTACGACAGCCAGCTCTCTCCATCCGAGGAGAAAGTATTGCTACAAAAGGAGTACTTT 209329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1092 AAGTCGAACGGTTTCTCTCAGGATGCAAACTGAAGTTGATTATGATTAGGACGTC 1151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209330 AGAGAGCAGAGCTCTCTCAAGAGCGCGTTCTCTCCGCAATGATTCAGGAACGAC 209389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1152 GTGGTTGCAAAACCCGAAAGTATCGATTAAACGAATTTGAAATTAATATAGACTCTCT 1211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209390 ATTATCAACTACAGCTGGAGATTAATCAATCAGAACTAGGAATCAATGTTGACTCCTT 209449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1212 CAGGAACGGAAAAAGATTAACCTCAGTGTGCCACAGCTCAGAAAGATATTGATAGA 1271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209450 AGCTTTAAGCAGCCCGTCAGCTTACAGCAAAAGTCTTCAAAATTAAGTATGATCTATC 209509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1272 TCGTCTGTGTACTGCGCAATTAAGGATGAGAGTTTATGCAAAATGCGTTTGAATGA 1331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209510 TGGGAAGCTCAACCTGATGATTAATGAAGGAACATTTATGAAGTCAATATGTTACAGCA 209569
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1332 GGACCATTCCT-----ATGATGGGATCTTGAGTTAGATGCTGG 1370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209570 TGACAGGCTCTCTCTCTATTAATAAATCAGCGTGTGATCTGATGTTACTAATCAGTTGA 209629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1371 GAAAGACATCGTATTTCTGAGATTCCTCGAGATATCAATAGCTGTACATCTCCGTATG 1430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209630 CATCAGCAGCTTATCCCTGTTCTCTGAGAGATCTTAATTCAGAAATGAGATTCGAAG 209689
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1431 CTAT---CAGGAAAGTGGACAAATCAATGTTGCTACTGATGATGAAGAAAGCTACGTTTC 1487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209690 ACAATGGAATGTTAATGAGTACAGGATACAGCTACAAATACAAAGAGCGACAGCAAC 209749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1488 TTGGGCAAGCAAGTTTAAATCCCACTGCTGAGCAGAGAGCTCGCTTATGTTCTAATCT 1547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209750 TTGGACCAAAACAGATATGTTGTTCCAGCCCGCAAGAAAGAAATCGCTTAGATCAAAATAC 209809
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1548 TCTTTGGGGTCTTTTATAGATGTTGCTGCTCCCAAAATTTTATAGAGTACAGTACGA 1607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209810 CTTATGGGAGCTTTTACTGACATTCGCTCTGCAACAGCTTTAGAGATGGGCGCAAC 209869
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1608 AGTGTCTCTTACGAAAGAGATTTGGGTTGACAGCAATTCATGTTTCAATGTTTGAATAGAG 1667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209870 TGTATGGAACACAACAAAGGTTTCTGGGTTCTCCATGACGAACCTCCGTCATTAAGAC 209929
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1668 CGGTCTGAAATCAAGGAATTCCTGATGTGAGTGGAGAGTGTGATGATAGTGTCTAG 1727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209930 TGGAGATCAAAATGCAAAAGGCTTCCTGTCATCTCTGGAGGCTACATCGATCGGTGGAG 209989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1728 CACGAGATGCGGGGTGTGTATCTGTCCTGAGGTTTCTGAGCTCTTGGGCGTGA 1787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209990 TGCTCACACTCTTAAAGACGACTATTTACTTTGCGTTTCTGCTTCTTCTGTAAGA 210049
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1788 CAAAGACTACTTTATGAATACCAATTTGCAAAAGCACTACGAGAGATCTTACGTTTGA 1847
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210050 CAAAGATGTTTATATGCTGTACCAACACTATAGAACCTACGAGTGAACCTTATATCTTCAA 210109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1848 GCAGAGTCTTCCCTATATCTGTGTGAGATATCTTTTATAGAGAGAGAGACTCCGGA 1907
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210110 GCATCTTCATACCTTACAAACCCCAAACTATTTAGATTAGGAAGAGCAAAAGTTTCTGA 210169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1908 GATCGTGTGCTTATGTTTCAAGACTCTGCGTCTCTTATGAGGACGCTTACGTA 1967
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210170 ATCAGCTATGAAAAAATTCCTTAGGAAATTCCTTACGCTTGGATGTGCAAGTTTGTT 210229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1968 CGGCATATGAGATCATGATGAAGACGAGTCTTACCCGCCGCCGCCGACGCTCTC 2027
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



COMMENT  
Mutsunori Shirai, Yamaguchi University School of Medicine,  
Department of Microbiology, 1-1-1 Minamikogushi, Ube, Yamaguchi  
755-8505, Japan (E-mail: mshirai@po.cc.yamaguchi-u.ac.jp,  
Tel: 81-836-22-2227, Fax: 81-836-22-2415)  
On Sep 15, 2000, this sequence version replaced  
gi:6172292 gi:6172294 gi:6172296 gi:6172326 gi:6172328 gi:6172330  
gi:6172332 gi:6172334 gi:6172336 gi:6172338 gi:6172382 gi:6172384  
gi:6172386 gi:6172388 gi:6172390 gi:6172392 gi:6172394 gi:6174666  
gi:6174668 gi:6635174 gi:6635176 gi:6635178 gi:6635180 gi:8547433  
gi:8547438 gi:8978640.  
AB033782-AB033785, AB033800-AB033815: Submitted (25-Oct-1999)  
AB038348-AB038349: Submitted (14-Feb-2000)  
AB036079-AB036082: Submitted (18-Dec-2000).  
Location/Qualifiers  
1. 300650  
/organism="Chlamydomonas pneumoniae J138"  
/strain="J138"  
/db\_xref="taxon:138677"  
/note="synonym: Chlamydia pneumoniae (strain J138)"  
144. .935  
/gene="CPj0267"  
144. .935  
/gene="CPj0267"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA98477.1"  
/db\_xref="gi:8978641"  
/translation="MSIMSKNTNMLNMPAPVCLANMPDKYINODKRTACTYTL  
VIATMITTGYIVYLLAMGSPRLSYLSTTIGTSTTGTALFTIGLKLKSLAMI  
OYKDEEVKOKTEPSPKNDVNHKLTPCLSPDLIESPSPASTVSKIRIACSG  
VAIVGVTLLGAVVSPFCGYQLALCVGACGTLALFVGGLGLRTHSLIAGIM  
YLTYLYLSALSEENETVKDQREINITYLTECHQKREKALLE"  
957. .1343  
/gene="CPj0268"  
957. .1343  
/gene="CPj0268"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA98478.1"  
/db\_xref="gi:8978642"  
/translation="MSQCSSTSTWEMKSFVPWKNPPLSPISPEDEFTIAYEP  
FVLPTDENAOANPGTSPVENGIDLNPDLQPNQONNANNPGTSGSNPISLP  
PERLPETEENSGEEOGSONNEDLIG"  
complement(1495. .2472)  
/gene="CPj0269"  
complement(1495. .2472)  
/gene="CPj0269"  
/codon\_start=1  
/transl\_table=11  
/product="dipeptidase"  
/protein\_id="BA98479.1"  
/db\_xref="gi:8978643"  
/translation="WTIDMHCDDLSPHPCRRKPAVRCSPQLLSGVGOVCALFVP  
HSRCPNCDKONSLEFPLNQYPDIGLISYEENSGSSOKKSLIRISINASLGD  
DTAGTGLTLAKLILHTKQGLAYLIGKGNRFGEGEAPKRLSNDGLDIYEL  
GVPTDLSHCSKRLADIDYADKLPNLAVTASHNFRSYVDHRRNLVDARAKETVR  
KGVIGLNLVRSYVDSGLDLEKHYHAENLIGLISIVLGSDFYVANEENPFNCCSS  
AAHPVYLNQHLIRIFSKKAKESILSSRAEKLKQIVQVNPKITDVKL"  
complement(2487. .3347)  
/gene="ywcA"  
complement(2487. .3347)  
/gene="ywcA"  
/codon\_start=1  
/transl\_table=11  
/product="Suas superfamily-related protein"  
/protein\_id="BA98480.1"  
/db\_xref="gi:8978644"  
/translation="MPDKAQTSLPEVMSAIHQKLTVALPTDVTYVGLVSLVSEA  
ERRVLRKREPKAFALVNSIEDIENISGPKLAOLEPGAITLVYKRN  
RPPKTLARIVDSVREIVDHGCTLIGTSANLSEPSALTAOEIPADPDHDCIT  
DPCSHGLESTVYASDPLKTYIREGLISVNIENAGTAKIFHRTSHAFSHKIKITYV

gene  
CDS  
3656. .4387  
/gene="CPj0271"  
3656. .4387  
/gene="CPj0271"  
/codon\_start=1  
/transl\_table=11  
/product="lysophospholipase esterase"  
/protein\_id="BA98481.1"  
/db\_xref="gi:8978645"  
/translation="WTDISFRRKIGTNEAIECPENPODPITILCHGYSIADNITTF  
PSICFSKLRPWPINPGLPLENDPFRSGRACFPNLVLLDELRLVANGNLOEKY  
DELFDVDETEPEKALEELILNLRPYNEIIGFSOGAILATHVLFQSNPYAGALIF  
AGARLFNOGMEGLKCAOVPELOSHEGDEELFPHGLAHLNDLLTLTNGCFVSFHG  
GHEIPSVYFQKQVVPNMIDPARG"  
complement(4359. .5255)  
/gene="anax.2"  
complement(4359. .5255)  
/gene="anax.2"  
complement(4359. .5255)  
/gene="anax.2"  
/codon\_start=1  
/transl\_table=11  
/product="DNA polymerase III gamma and tau"  
/protein\_id="BA98482.1"  
/db\_xref="gi:8978646"  
/translation="MHLEENOGWEALRKVYHOEVPAILNGFTLPVLODKAEOLA  
SEILSSPGEHKKVSKIHPIQVFPPEGKRLSIDLPRIKQIYISPEANVKI  
YIIEADMTYLAISAFKVEEPKHAVIITLTQKRLPRTIISRSISYIEGK  
TICSEKPSYLFRAVACLPTEVSOITKESEDDKQVLRKVOFPMVLEIETDRY  
TINLGLKASALNPYHVKETIQPLPLPDKVILLIYESACRLSNSSASVILEVAIO  
LVSLQYKRELVSVSPGODLSN"  
complement(5246. .5866)  
/gene="tdk"  
complement(5246. .5866)  
/gene="tdk"  
/codon\_start=1  
/transl\_table=11  
/product="thymidylate kinase"  
/protein\_id="BA98483.1"  
/db\_xref="gi:8978647"  
/translation="MEIVYEGEGSGSKSLAKIGDOLVADNRKVLIRPQCGCLICE  
RLRDLIRPHELSRCCELFPLDSRQHOEVIITPALRQGIYICGRHDSITVYO  
GIAEGLADPEVADLCSKVAGPPPLPNVLLDIPADIGLQKHKQKVFDEKPKPLS  
YHNRIRGELFSLASDPRLVPLDRESLASLIDKVMLTQGLCT"  
complement(5871. .8375)  
/gene="gyra.1"  
complement(5871. .8375)  
/gene="gyra.1"  
complement(5871. .8375)  
/gene="gyra.1"  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase subunit A"  
/protein\_id="BA98484.1"  
/db\_xref="gi:8978648"  
/translation="WFKNDEITVFNKLEEMKESYLRSMYSIIRALPDTRDGLKPS  
QRRVLYAKMQLSLSPGAKHRCCKITCGTSDYHGHGSVITYPLVRAQWAMVYPL  
VDGQGNESIGDDPPAARVYEARLTSHAMLEMDLDDVDYVNPYBETHEVVPV  
SKFPLNLGNSGSLVAGMATNIPHNLELMEATLLILANQASVDELTQWMPGDFP  
TGGITLCSGESSYRTTGRKIKVARRHVENENKESYIITPMPNVKNRSLIEO  
IANLVNKTLAGISVPRDSQDGRVLETKIGGSSITINRLYKFTDVOYTRANM  
IALDKNLRPTMSIRHMSAMTRHREKVIIRRTIRLELNKAEFRARHLESLALSLDA  
LVKTIRESGNKEHAKERIIIESGFTEPDALILRLYQLTGLEAKTQKEEELNK  
IAYKQVLSDEGLVNDIIRNELODLKHKKARRTTIEPDADIDIDIDITNESVLI  
TISGDVYKRMVYKFEORRGHGVTFDMKKGGLKAVYSAFTKDYLLFTNFGQ  
CYMLKVMQLPEGERAKGPIINFLGIRPEELAAIINKFNDAAGFLFATKRGV  
KVSILDAFSPNRKGIIRALEIDEGDELIAACHIVSDEKVMLEFTHLGAIVRPHKVR  
PMGRARGVRSYSLKNEKRYVSCQIYENSSVLYICOGGKRSYVADREPRTNGV  
GVRSTLINRGNVIGALIVTDHDSILMSSOGQAIRINMODVRYMGSTGCVRLVHL  
REGALVSMKLSNENDEVLSSGSECSPTLSR"  
complement(8391. .10808)  
/gene="gyrB.1"  
complement(8391. .10808)  
/gene="gyrB.1"  
/codon\_start=1



QY	796	--CATAGTGTATTTGATTTAGGAACACTAGCCCTAAGATTAACCTGCTTACGTCGCGCA	851
Db	12578	TAAAGCTAAATGGGATCATACATATAGGAAGTACGGGAATTCACGGGAATTCGGGCGTGCCTAA	12637
QY	852	AGGACATACGATATATCTTTTATGATCCGATTTACTGTATACAGGATGCACATCTGTGGCTGA	911
Db	12638	AAATCATCAATTTTCTTCTATGATGCCATCTTACAGAGGAACTCTAT-----CAGA	12691
QY	912	TGCTCTCAATATTTAATAGCCCTGATACCTGAGATTAACAAGATATACGGGACATAGCT	971
Db	12692	CGTATTTGAAGATAAATTAACGGCTCTCGGAGACTCTCAATCCATATCAAGAAACGATTTCT	12751
QY	972	CTTTTCTGGAGAGAAAGCTACGAGGACGAGAACTAAAGATGAGAAAGACCGCACTTCTTA	1031
Db	12752	ATTTTCTGGAGAAACCCATACAGCAGATGAATACCTTTAAAGTTGGTGCACAAATTTTAAATCTTTC	12811
QY	1032	ATTACTTCAAAATGTGCTTTTAAAAATGGGAGCTGATGTTTAAAAAGGTGATGTCGTTT	1091
Db	12812	ATTTACAGCAGCCAGTCTCCCTATCCGGAGAAAATTATTGCTATCAAAAGGAGTCACTTT	12871
QY	1092	AAAGTGGAAAGGTTTCTCTCAGAGATGCAAACTCTAAGTTGATTATGATTTAGGAGCGTC	1151
Db	12872	ACAGAGACAGACTTCTCTCAAGAGCCGGTCTCTCCGCGGATGAGATTACAGAAACGAC	12931
QY	1152	GTTGGTTGCAAAACCCGAAAGATGATGAATTAAAGAAATTTGAGAAATTAATATAGACTCTCT	1211
Db	12932	ATTATCACTACAGCTGGGAGATTTACATATCAGAACCTAGGAATCAATAGTGCCTCTT	12991
QY	1212	CAGGAACGGGAAAAAGATTAACCTAGTCTCCACAGCTGCAGAAAGATATTCGATATGA	1271
Db	12992	AGGCTTTAAGCACCCGCTAGCTTAACACCAAAAGGTGCTTCAAAATAAAGTATGCTATAC	13051
QY	1272	TGCTCTGTTGTACTGCGCAATTAAGGATGAGATGTTTATCAAAATGGCTTTTGAATGA	1331
Db	13052	TGGGAAGCTCAACCTGATGATTAATGAAGGAAACAATTAATGAAGTCATATGTCAGGCA	13111
QY	1332	GGACCAATTCCT-----ATGATGGGANTCTTGACTTATGATGCTGG	1370
Db	13112	TACACAGCTCTTCTCTCTATTAATAAACACGGTTGATGCTGATGTAATCACTAACGTTGA	13171
QY	1371	GAAAGCATCGATGATTTCTGCGAGATTTCTGCGATTAATATGATGTATACATATCCGATAGG	1430
Db	13172	CATCAGCAGCCTTAATCCTGTCCTGCTGATGAGATCCTAATTTAGAAATGCGATTCACAAAG	13231
QY	1431	CTAT--CAGGGAAGTGGACAACTAATTTGTCTACTGATGTAAAGAAACCTACGCTTTC	1487
Db	13232	ACAAATGGAATGTTAATTTGGACTACGGAATACACTCAAAATACAAAAGAGCCACGGCAAC	13291
QY	1488	TTGGGCAAAAGCAAACTTTAATCCACATCTGACAGAGAGGCTCCGTTATGTTCTAATCT	1547
Db	13292	TTTGGACCAAAACAGATTTGTTTCCACAGCCCCCAAAAGAAATCTGGTGTAGATATGCAATAC	13351
QY	1548	TCTTTGGGGTCTTTTATATGATGTTGCTGCTCCCAAAATTTATAGAGCTATGATACGA	1607
Db	13352	CCTATTTGGGAGCTTTTACTGACATTTGCTCTTGCAACGCTTGATGAGANTGGGCGCAAC	13411
QY	1608	AGGCTCTCTTACGAAAAGAGATTTTGGGTTGCAAGCAATTTCCAATGTTTGCATAGAG	1667
Db	13412	TGGTATGGAGACAAACAAGATTTCTGGGTTTCTTCCACAGACGAACCTTCTGCAATAGAC	13471
QY	1668	CGGTCGTGAAAAACAAGAAATTCGTCATGTGATGAGTGGAGGCTGTATAGTGGTCTAG	1727
Db	13472	TGGAGATGAATAATGCCAAAGGCTTCGTCATACCTCTGGAGGCTACGTCATGCGTGGGGAAG	13531
QY	1728	CACGAGGATGCCGGGTGGGATACCTTGTGCTGGGTTTGTCAAGCTTTTGGCGCGTGA	1787
Db	13532	TGCTCAGACTCTTAAAGACGACTATTTAATCCTTTGCGTTCTGCAATCTCTTGTGCTAGAA	13591
QY	1788	CAAAAGATCTTATGAATATCAATTTTGCAAAAGACCTACGAGAGATCTTTACGTTTGA	1847
Db	13592	CAAAATTTGTTTATTCGCTCACACAACTCTTAGAAACCTACAGGTGGAACTTTTATTTCTCAA	13651
QY	1848	GCACGATCTTCCCTATCTCTGTGTGATGATATCCTTTAGGAGAGGAGAGACTCCGCA	1907

DB	LOCUS	DEFINITION	ACCESSION	VERSION	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL
Db	13652	GCAGCTCTACACCTACACACCCCAAACTATTGATTTAGATTAGAGACGCAAACTTTCTCGA	13711												
Qy	1908	GATCCTGTGGCTTATATGTTTCCAGACTCTGCGGTCTCTTCTATATGGGACGCTTAGCTA	1967												
Db	13712	ATACGCTATAGAAAATTTCCCTAGAGGAATTTCCCTACGCTTGGATGTCCAAAGTTGGTT	13771												
Qy	1968	CGGCAATACGGAATCATCCGATGAGACGAGAGTCTCTACCCCCCCCCCGGACGCTCTC	2027												
Db	13772	CAGCAATTCAGCAACCCATATGGAACCCACTATACCTCATTTGCGAGATC-----	13822												
Qy	2028	GACGATCATCTACTTCTTGGGAGGATATGTCTGGGCTGAGAGGCTGGGAACCTCGAGTTGC	2087												
Db	13823	-----CGAGGTTCTTGGAGCACGAGATGTATACTTGTGTGTATCGGCTAGACCTTCC	13876												
Qy	2088	TGTTGAAAATACACAGCGGACAGAGATTTTCTCGAGAGTACATCCATTTGTMAAGTCCA	2147												
Db	13877	TTTTGTCTCTTCCAAACCCACATCCCTTTTCAAGACCTTATTCACAGATGAAAGTCTGA	13936												
Qy	2148	AGCTGTTTACTCGCGCCAGATAGCTTTGTTGAACTAGAGACCTATCACTGTATTTTAC	2207												
Db	13937	AATGTTTATGATACAAATATAGCTCTTTCGAAAGCTCTAGTATGATGAGCCCTGGTTTAC	13996												
Qy	2208	TCATTCGATCTTTTATACCTTGGCATTCCTCTTGGATTCAGTATGAGAAACGCTTTC	2287												
Db	13997	TATTTGAAGCGCTTCAACCTCTCAATTCCTGAGGTCGAAATTCGCGAGGGGATAT	14056												
Qy	2268	AGAGCAATA-----TTATACATGTTGTAGCGATGATTTCTCAGAGTGTGTTTCGTAGTAA	2321												
Db	14057	CGGAGATTCCTTACACCTATGATCTCTCAGAGATCTTCTTTCGATGTCTATTCGTAACTA	14116												
Qy	2322	CCCCAATGTACGACTACCTCATTTCCACACCAAGGAGTTGAGAACCAAAAGTTGCA	2381												
Db	14117	TCGCCATCTACAGGACTCTGTGATGAGCCACAGACTCTTGAAAAATTTCCGCTGGCAA	14176												
Qy	2382	CTTAGCAGACAGCGCTGATTTGTTCAGGCTCAGAGTTTGATCTTTGGAGCTGCAGC	2441												
Db	14177	TCCTTCAGACAGGACTTTTACTGAGGGGTAGACAACTAGCTCTAACCTCAATTCG	14236												
Qy	2442	AGAGCTTTTGGGGAATTTGGCTTTGATATGGCGGAGCTCTCTCGATATATATGAGA	2501												
Db	14237	TCAGCTCTTCGACACTTACGCTATGAGAACTCCGTGATCTTCAAGAACTCATATGTAGA	14296												
Qy	2502	TGCGGGTAGCAAAATCAAAATTTTAC	2526												
Db	14297	TGTTGTATCCAAACTCCGATTTCTAG	14321												
RESULT	12														
AP002546	300650 bp	DNA	BCT	08-DEC-2000											
LOCUS	Chlamydomophila pneumoniae J138	genomic DNA, complete sequence,													
DEFINITION	section 2/4.														
ACCESSION	AP002546	AB033782	AB033783	AB033784	AB033785	AB033800	AB033801	AB033802	AB033803	AB033804	AB033805	AB033806	AB033807	AB033808	AB033809
	AB033810	AB033811	AB033812	AB033813	AB033814	AB033815	AB033816	AB033817	AB033818	AB033819	AB033820	AB033821	AB033822	AB033823	AB033824
	AB033825	AB033826	AB033827	AB033828	AB033829	AB033830	AB033831	AB033832							



QY 1728 CACGAGATGCCGGGTGGATACCTGTCTGGGTTTGTGCTACGCTTTGGCCGCTGA 1787  
 Db 2866 TGCCTACACCTCTTAAGACGACCTATTACCTTTGCGTCTGCGACCTCTTGGCAGGA 2807  
 QY 1788 CAAAGACTACTTTATGAATACCAATTTGCAAGAACCTTACGACGATCTTTACGTTTGA 1847  
 Db 2806 CAAAGATTTGTTTATCGCTACCAACACTGAGAACCTACGCTGAGAACTTATCTTCAA 2747  
 QY 1848 GCAGATGCTTCCCTATCTGTGTGATGATCTTTTAGAGAGGAGACGCCGGA 1907  
 Db 2746 GCACCTCATACCTACCAACCCCAAACTATTGATGAGAGAGCAAAAGTTTCTGA 2687  
 QY 1908 GATCCGTGCTTATCTTTTCCAGACTCTGCGGTCTTCTTCTATGCGACCTAGCTA 1967  
 Db 2686 ATTCAGCTTATCAAAAATTCCTTAGGAAATTCCTTACCTTGATGTCAGATTTGCTT 2627  
 QY 1968 CGCCCATACGATCATGCGATGAAAGCCGAGTCTCTACCCCCCCCCCGACGCTCTC 2027  
 Db 2626 CAGCATTCAAGACAACCTATGAAACGCACTATACCTCATGCGCAGAAATC----- 2574  
 QY 2028 GACGATCATCTTCTTGGGAGAGATATGCTGGGCTGAGAGAGCTGGAACTCGAGTTGC 2087  
 Db 2575 -----CGAAGGTTCTGTGAGACAACGAGTGTATAGCTGTGGTATCGCTAGACCTTCC 2522  
 QY 2088 TGTGAAAAATACACGCGCAGAGAGATTTTTCGAGAGTACACTCCATTTGTAAGTCCA 2147  
 Db 2521 TTTTGTCTTTCCTCAACCCACATCCCTTTTCAAGACCTTATTCACAGATGAAGAATGGA 2462  
 QY 2148 AGCTGTTTACTGCGGCCAAGATAGCTTTGTTGAACCTAGAGACCTACCTGCTGATTTAG 2207  
 Db 2461 AATGTTTATGTATACACAAATACCTCTTCGAAAGCTCTAGTGTGCGCTGCTGTTTGA 2402  
 QY 2208 TGATTCGATCTTATATACCTTGGATTCCTCTGAGATCAATCAATGAGAAAGCTTTCG 2267  
 Db 2401 TATTGGAAGCTGCTTATACCTCTGATTCCTGTGGGTCGAAATTCGTGCGAGGGATAT 2342  
 QY 2268 AGACCAATA-----TATCATGTTGTAGCGATGATTCCTCAGATGTTTCTGCTAGTAA 2321  
 Db 2341 CGGAGATTTCCACACACTATCTCTCAGATTCCTTTGTTCCGATGTCATCGTAACAA 2282  
 QY 2322 CCCCAATGTACGACCTTCTTCCACCAAGGAGTTGGAAGACCAAAAGCTTTCGAA 2381  
 Db 2281 TCCCAATCTACGACGCTCTGTGATGAGCCAGACCTTGGAAATTCGCGGTGCGAA 2222  
 QY 2382 CTTCAGCAAGAGGCTGATTTGTTCAAGGCTCAGGTTTGTGATCTTGGAGCTGCGAGC 2441  
 Db 2221 TCTTCAAGACAGGCAATTTTACTAGGGGTAGCAACAACACTCTACCACTCAATTTG 2162  
 QY 2442 AGAGCTTTTGGGAACCTTTGGCTTTGAATGGCGGGGATCTTCTGTAAGCTTAATGTAGA 2501  
 Db 2161 TGACCTCTTGGCAATTAAGCTATGGAACCTCGTGGATCTTCAAGGAACCTACATGTAGA 2102  
 QY 2502 TGGCGGTACCAAAATTCATTTTGA 2526  
 Db 2101 TGTGTGTAACCAACTCGATTTCTAG 2077  
 RESULT 11  
 CPN133034 17280 bp DNA BCT 27-FEB-2001  
 LOCUS CPN133034  
 DEFINITION Chlamydia pneumoniae ggal-protein family cluster A, genes omp12  
 (partial), omp11, omp10, omp5, omp4, omp13 and omp14 (partial) and  
 OmpL, strain VR1310.  
 ACCESSION AJ133034  
 VERSION AJ133034.1 GI:4455881  
 KEYWORDS hypothetical protein; omp10 gene; omp11 gene; omp12 gene; omp13  
 gene; omp14 gene; omp4 gene; omp5 gene; OmpL; outer membrane  
 protein; outer membrane protein 11; outer membrane protein 12;  
 outer membrane protein 14; outer membrane protein 4; outer membrane  
 protein 5.  
 SOURCE Chlamydia pneumoniae.  
 ORGANISM Chlamydia pneumoniae  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE 1. (bases 1 to 17280)  
 AUTHORS Daugaard, L., Hiern, K., Knudsen, K., Madsen, A.S., Christiansen, G.  
 and Birkeland, S.  
 JOURNAL Unpublished  
 REFERENCE 2. (bases 1 to 17280)  
 AUTHORS Boesen, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-JUN-1999) Boesen T., Department of Medical  
 Microbiology and Immunology, University of Aarhus, The Bartholin  
 Building, DK-8000 Aarhus, DENMARK  
 FEATURES  
 source  
 1. 17280  
 /organism="Chlamydia pneumoniae"  
 /strain="VR1310"  
 /db\_xref="taxon:83558"  
 1. 841  
 /gene="omp12"  
 /note="omp12"  
 /gene="omp12"  
 /note="putative"  
 /codon\_start=2  
 /product="outer membrane protein 12"  
 /protein\_id="CAB37067.1"  
 /db\_xref="GI:4455882"  
 /db\_xref="SPTREMBL:O92504"  
 /translation="EDNIRYRHHNGGVALGTTATPAEDQLTFAPCOLFARDRHNTG  
 KKHGDYGLASLTFHHTEGLFDIANFLMGKTRAPWMLSEISQIIPISFPAKESYLFED  
 NHHKTYTDSITIKGSWRNDAFCADLASLIPYISVPLYLEKEPEPVKQYIYAHOD  
 FYERHAEGRAFNKSLEINDEIPGVFEERDSKSEKQYDITLWYIILDAYARHCOTS  
 LIASDNNMAYVGNILARQGFVRAAHNFQVNPHEIFQGFAPFVRESSSRVNTNLSK  
 FCR"  
 885..967  
 /gene="omp11"  
 885..3677  
 /gene="omp11"  
 885..3677  
 /gene="omp11"  
 /note="putative"  
 /codon\_start=1  
 /product="outer membrane protein 11"  
 /protein\_id="CAB37068.1"  
 /db\_xref="GI:4455883"  
 /db\_xref="SPTREMBL:O92393"  
 /translation="MKTIPLKHLISLTVYPIILLSTATYCADASISPTSPDAGST  
 FTRKSTADANGTVYLSGNYINDAGGRLITCCFTETTGDTLFTGKTSREFTND  
 AGSNAGAASATTDKALTTFTGFNSLFTAPGTTVASGKSTSSAGALNTDGTTLF  
 SQWVSEANNGAATYTKTLISIGNTSITFTNSAKKLGAGIYSSAASISGNTQL  
 VFNNGGTEGCGALGFASSTIONSLSFSGWTADAGKGAICYERTGETPILTI  
 SGKSLTPAENSSTVTOGGAICAHGLDLASAAPTULFNNRCGNTAAKGGAIAADSGS  
 LLSANAGOTTPILGNLTFSAPTSTRNATYLESSAKITNLRAGOSIYFDPDASN  
 TTGASDVLITNOPDSNPIDYSGTIVFSGEKLSDAKKANDNTSLKPLAAGSTL  
 ALKGNVELDVNGFTQEGSTLLMQPETKLADEALSLKLVVDLSALGKNSVSLET  
 AGANKTITLTSPLVFDSSGNFESHITINDAFQPLVFTFAAASDIYIDLTLSPV  
 QTEPEHYOGHEATWADSTAKSGTMTVTTGYNPDRASVGYIVGASFTDIR  
 TLGOIMTSOANSIYOOGRGLASGTANFPHKDSQGNAPRHSYGYIVGASDESEN  
 IESVACOLFGKDKDLFIYENTSHTNLASLYLOHRAPIGLGIPRSGSITDMIKDIL  
 ILNAQLSYSTRKNDTRITSTYERAGSGTNNSGALELGSALYIIPKAPFQGTGP  
 FLKFOAVYSRQNFKESSGAERAFDGDVLVNCISIPVIRLEKISEDEKNNFEISLAVI  
 GDVYRNRPSTSLWVSGASWTSICNLARQAFVLAGSHLTLSPHVELSGEAAVELR  
 GSAHIVNDCGLTSF"  
 3827..3907  
 /gene="omp10"  
 3827..6613  
 /gene="omp10"  
 3827..6613  
 /gene="omp10"  
 /note="putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="outer membrane protein"  
 /protein\_id="CAB37069.1"



Db 10515 TGAGCTCTCGACATTACGCTATGACCTCGTGAATCTTCAAGCAACTACATGTAGA 10574  
OY 2502 TGGGGTAGCAAAATCAATTGTAG 2526  
Db 10575 TGTGTGACCAACTCCGATTTCTAG 10599

RESULT 10  
AE002192/c  
LOCUS  
DEFINITION  
Chlamydia pneumoniae AR39, section 25 of 94 of the complete genome.  
ACCESSION  
AE002192 AE002161  
VERSION  
AE002192.2 GI:8163405  
KEYWORDS  
ORGANISM  
Chlamydia pneumoniae AR39.  
SOURCE  
Chlamydia pneumoniae AR39.  
REFERENCE  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
AUTHORS  
1 (bases 1 to 12676)  
Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,  
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,  
Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,  
Bowman, C., Dodson, R., Gwin, M., Nelson, M., DeBoy, R., Kolonay, J.,  
McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.  
Genome sequences of Chlamydia trachomatis Mohn and Chlamydia  
pneumoniae AR39  
Nucleic Acids Res. 28 (6), 1397-1406 (2000)

TITLE  
2 (bases 1 to 12676)  
Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,  
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,  
Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,  
Bowman, C., Dodson, R., Gwin, M., Nelson, M., DeBoy, R., Kolonay, J.,  
McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.  
Direct Submission  
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
On Jun 1, 2000 this sequence version replaced gi:7189226.  
location/Qualifiers

COMMENT  
FEATURES  
source  
1..12676  
/organism="Chlamydia pneumoniae AR39"  
/strain="AR39"  
/db\_xref="taxon:115711"  
/note="synonym: Chlamydia pneumoniae AR39"  
gene  
complement(90..185)  
/gene="CP0300"  
complement(90..185)  
/gene="CP0300"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
CDS  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AF38157.1"  
/db\_xref="GI:7189227"  
/translation="MFTSLGRSAIFFTGSLKHLRGEFLIRQK"  
complement(275..1819)  
/gene="CP0301"  
complement(275..1819)  
/gene="CP0301"  
/note="similar to GP:4376735; identified by sequence  
similarity; putative"  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="polymorphic membrane protein G family"  
/protein\_id="AF38158.1"  
/db\_xref="GI:7189228"

gene  
CDS  
complement(2077..4926)  
/gene="CP0302"  
complement(2077..4926)  
/gene="CP0302"  
/note="similar to GP:4376733; identified by sequence  
similarity; putative"  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="polymorphic membrane protein G family"  
/protein\_id="AF38159.1"  
/db\_xref="GI:7189229"

KNLRVSESGVYFYPDIISHSEHKITDVLINAEKGEYEGTISFSGCLUDHENCAL  
NLRTSLIDVYTLAGTGLSLDGVTLQIOLHSEKQASSTLIPAGTPTLLIGSGDAVOMH  
ILTEPDNDPVPVIRAPEDDAIVSLERKIVAFAPASVYDFDPKRAPIPLLEILGP  
SEPSLLIGETTLERTOYTTENDAVRQFMSLMEYEPSPDLKDRITPTKRYVLTWNP  
ELITSip"  
complement(2077..4926)  
/gene="CP0302"  
complement(2077..4926)  
/gene="CP0302"  
/note="similar to GP:4376733; identified by sequence  
similarity; putative"  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="polymorphic membrane protein G family"  
/protein\_id="AF38159.1"  
/db\_xref="GI:7189229"

gene  
CDS  
complement(8170..8292)  
/gene="CP0304"  
complement(8170..8292)  
/gene="CP0304"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AF38161.1"  
/db\_xref="GI:7189231"  
/translation="MDLSRWMLNKLKTRQOLDQNIKDYDCKSYCLRGISQILGMV"





QY	2088	TGSGAANAATACAGGGCGAGAGATTTTTCCAGAGTCACTCCATTTGTAAAGTCCA	2147
Db	5247	TTTTTGTCTTCTTCAACCCACACCTCTTTTCAAGACTTCATTTCAACAGATGAAGGTGA	5308
QY	2148	AGCTGTTTACTCGGCCACAMATAGACTTTGTGTGAACTAGGAGGTATACAGTCGATTTTGA	2207
Db	5307	AATGGTTTATGTATCAACAANAATAGCTTCTTCGAAACCTCTAGTGAATGGCGCGGTTTGA	5366
QY	2208	TGATTCGCATCTTATTAACCTTTCGCATTTCTCTCTGGAAATCAAGTTAGAGAAACGTTTGC	2267
Db	5367	TATTGGAAAGCGCTGTATTAACTCTTCGATTTCTCTGGGTGCGAAATTCGTGCAGGGGATAT	5428
QY	2268	AGAGCAATA-----TATCATGTCTTAGAGATGATTTCTCAGATGTTTGCAGTAA	2321
Db	5427	CGAGATTTCTCAGACCATATATCTCTCAGAGATTTCTTTTCCGATGTCTATCTGTAACAA	5488
QY	2322	CCCCAAATGTACGACTACCCCTACTTTTCAACAACGAGGATGGAAGAACCAAGTTTGAA	2381
Db	5487	TCCCAATCTACAGCCGACTCTTGTGATGAGCCGACCTTGTGAAAATTCGGGGGCGCAA	5548
QY	2382	CTTAGCAAGACAGGCTGTATTTGTTCAGGCGTCAGTTTTCATCTTTGGAGCTGCAGC	2441
Db	5547	TCTTTCAAGACAGGCAATTTTACTGTGAGGGGTAGCAACAACATAGCTCAACTCCAAATG	5606
QY	2442	AGAGCTTTTGGGAACTTTGGCTTAAATGAGGGGAGCTTCTCTAGCTATATGTAGA	2501
Db	5607	TGACCTCTTGGACATTAAGCCTATGGAACATCGGTGATCTTCAAGGAACCTACAAATGTAGA	5666
QY	2502	TGCGGGTAGCAANAATCAATTTTGA	2526
Db	5667	TGTTGGTACCAAACTCGATTTGAG	5691

RESULT	9
LOCUS	AEO01628
DEFINITION	Chlamydia pneumoniae section 44 of 103 of the complete genome.
ACCESSION	AEO01628 AEO01363
VERSION	AEO01628.1 GI:4376730
KEYWORDS	
SOURCE	
ORGANISM	Chlamydomydia pneumoniae CWL029. Chlamydophila pneumoniae CWL029 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomydia.
AUTHORS	1 (bases 1 to 10757) Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W., Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S. Comparative genomes of Chlamydia pneumoniae and C. trachomatis Nat. genet. 21 (4), 385-389 (1999)
TITLE	2 (bases 1 to 10757) Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S. Direct Submission
JOURNAL	Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
MEDLINE	
PUBMED	
REFERENCES	
FEATURES	location/qualifiers 1..10757 /organism="Chlamydomydia pneumoniae CWL029" /strain="CWL029" /db_xref="taxon:115713" 106...2892 /gene="pmp_9" /note="CPn0447" 106...2892 /gene="pmp_9" /codon_start=1 /transl_table=1 /product="Polymorphic Outer Membrane Protein G/I Family" /protein_id="AADI859I.1" /db_xref="GI:4376731" /translation="MKSSIHMFLLSSSLAPLSLNFSAAVAVEINLCPTMSFGPGTG"

BASE COUNT	ORIGIN
3228 a 2459 c 2065 g 3005 t	<p>             YTPAQTINMADGTYINLGDVSTINMAGSPALTLTSCRETTGNTLSPGCHGCPHLOLNT              DAGNAGTFTNTAANKLTLFSFGSTLSTLQTTNATGATGALSTAGCSIQNSTCYGQF              NESNDNGALGAGSSISLSLNLNPLFAKKAAYOKGALYLTGGITINNTLNLSASESNT              AANNGALTYEASFSISFNKAISFISNNSVTATGCIYCSYSPARPVLTLDNGL              LNFNGNTATGSGALITPDNLVLSSGPIFLFNKNSAIATGASAPALADGSLTLDNGL              LGGITPFGTNGVVGASSQTTTNSIGNTKNTIYOLRASQNTIYFVDPITTSITL              AALSDALNNGPDLAAGNAYGCTIYFSEKSEKPAEADMLKSTIOOPLTLAGGOLS              LKSGVTLVAKFSQSPGSLTMDAGTLETETADGITTINLVNLDSLKETTKATIKATQ              ASIHQVYQSSLSLVDPSGVNVEDYSWNNPOVSCLTLTADDPANIHTDLADLPEK              PIHNGYUOGNMLSMQEDTATSKATLTMTATGATNPNPERGATLVANTLWMSFVDRK              IQOIVATYVROSQETRCIWMCEGISNFEPIKDSCKINCKGRALISAGYVGAATLTASND              ITAPACOLFGKDRDHFINKNRSAYASLHLOHATLSTPSPLLVLGSESEOPVLEF              AQISYISKNTMYTYQAPKGESSWYNDGALALASSLPATLSHGCIAPAYPEPKIK              VEASTIHODSFKERKNTLVRSFDSGDILNVSPIGITIFERSNRERASTETATVYAA              VYRRNPDCITALLINNTSMKTTGNTLSHQAGIGRAGITFASPNLEVTSLNLSMEIRRS              SRSYNADLGKRFQF"              complement(1233, .4348)              /gene="yyj6_2"              /note="Cp00448"              complement(323, .4348)              /gene="yyj6_2"              /codon_start=1              /transl_table=1              /product="hypothetical protein"              /protein_id="AADI8592.1"              /db_xref="gi:4376732"              /translation="MEMSPFOQPEOCHFVYVGSFLRPESLTRARSDFEGRGIYEQM              RLVDAIARLNLIKKTQTEAGLIFFTDGEFRARYSNQPEDMWGPHGVDRSDNDPIGVY              LKIDISYSKHPETHEFEFFKFEKGNKAKOTIPSPQOFHEMIFAPLNKTRKFEYPLP              NOELIDIVYRYROYIDLYAACGRNQDLDCAMCRLLDIRAPMYGVDSIDRCQETLIL              EQLFNIHLWKKDRPEDLVSLHVCRDQYAEFSSRATVDEEPLAKTDVDSITHY              WALDDKYSGCAEPLAYVSGEKHVCGLTISNHSQIEORDAVSRIVYEAASYIPLERLS              LSPQCFASCEGDRMTEDQMKKIAEVKEIAKEIWG"              complement(4655, .6769)              /gene="pmp_10.1"              /note="Cp00449"              complement(4855, .7650)              /note="Polymorphic Outer Membrane Protein G Family;              contains frame shift"              complement(6715, .7650)              /gene="pmp_10.2"              /note="Cp00450"              7813, .10599              /gene="pmp_11"              /note="Cp00451"              7813, .10599              /gene="pmp_11"              /codon_start=1              /transl_table=1              /product="Polymorphic Outer Membrane Protein G Family"              /protein_id="AADI8593.1"              /db_xref="gi:4376733"              /translation="MKTSLPWVLYSSVLAESCHIOSLANEELTSPDPSFNGNDGTF              TPKTSATYISLIGDVFYREPEKGTPLBDSCEPKQTTDNLUTPLGNGHSLTFEGTIDAGTARA              GAASATANKNLTLFSGFSLTSDSPSTVYTGQTLSSAGVULENIRKLVVAGNS              TADGAIKFGASFLTGTSGDALFENNSSTGGAIAATGAGRLANNNGYVAFLSNFS              TSGAIDDEGTSILSNKKPLFEEGNAAKTGGAICNTRKASGPELILSNKNTLFAASN              VAFISGGAIAAKKIALSSGGTFEELRANNSVAPKGAISIDAGELSLSAENITFET              VRNLTITGTSIDTPYKRNAINGSGKRTPELAKNHTTFPDDPTTSCGTSQVLIKINN              GSAALNPYQGTILFSEETLADLAKVADNKSFTOPVLSGGKLLQKQTYLESTYS              FSGAGSLTMDSGTTLSTAGSITITNLGINVDSIDLKQVSLTLAGASAKRYVASKR              LNLIDEGNITYESHMFSDOLFSLKLTIVADVDITNDISLIPVAPEDNSEYFQGGAG              OMANNWTIDTATNTKEATATWTKGFPVSPERKSAVCNTLMGVFTDIRLQOLVELTIG              ATGGEHQEFVWSSMTNPLKHTGDENKRGPHRSGGVVIGSAGTPKDDLETTFPACHL              PARKKDFEIAHNNSTRTYGTILFFKHSHTLOPNTLRIGRAKFSASALEKPREPLAL              DVQYFSDNDNRNMTHTYSLSPESGSSNCEIAGIGLIDLPVLSNHPLEKTIPOM              KVEVAYVONSFEVSSSDGRGFSIGRLNLNIPGARGFVQGDIDSYTLDLSGFFVSD              VYRRNPSTALIVSSPDSMKIRGNLSRQAFLLGSSNNVYVNSNCELFCHYAMLRSS              SRRNVVGVGTRFPR"  </p>



```

/translalion="MKTSLIPVIVSSVLAFSCHLSLANEELLSPDSEFNGIDSGT
TPKTSATYSLTGVDFEYEPKGFPLSDSCFPCOTDNLFLFNGSLFEGFDAGTHA
GAASTTANKNTLTFSGEFLSPDSSPTTWTTCOTDNLFLFNGSLFEGFDAGTHA
TRDGAIRKASFLITGSDALFNNSSSTKGAATATGABIANNTGVRLSINAS
TSGGAIDEGTISILSNKFLYEGNAATGTGALNTASGPBELLISNNKTLIPASN
VAETSGGAIIHAKKALSSGTEFLRNWVSATPGGAIISIDASELISAFEGNITF
VANTLTGTSSTDPKRNAINIGSNKFLPDLAKHNTLFFYPITSEGSVYLKLN
GSAALNPOGTILFSGEFLPADLKNLKSFTOPVLSGKLLLOKGYLEST
FEOEAGSLIGMSGTTLSITAGSITITNGIDVNDLSLIPADPNSEYFSG
LNLIDENIVSEHSHDLSLILKIDVNDLSLIPADPNSEYFSG
OINWMTTIDTATNTKEATATATKTFVSPERKSLVCTMLGVTIDRLSQVLK
ATGEMHKGFEWVSMTNFKLTKDNRKGFRTSGGYIIGSAHPKODLSAFEGNITF
FARPDCEFIANNSTRTGTLFFKSHSLIOPONYRLGAKSESABIEKFEFIDAL
DVQVFSHSDNMEHTHSLPESSEGSMECIAGIGLDPVLSNPHLFTFPIOM
KVENYVSONSEFESSDGRGSIIGRLNLSIPVAKFIQODIGDSTYTDLSGFEVSD
VYRNPOSTATILVNSPDSMKIRGSLNSQAFILRKSNNVYVNSNELGHYAMELRGS
SRNYVDYGTKLRF"
mat_peptide
2956..3688
/gene="omp4"
/product="outer membrane protein 4"
stem_loop
5698..5726
/gene="omp4"
/feature="putative"
BASE COUNT 1760 a 1341 c 1195 g 1734 t
ORIGIN

```

```

Query Match 13.5%; Score 341; DB 2; Length 6030;
Best Local Similarity 50.1%; Pred. No. 6.5e-72;
Matches 1095; Conservative 0; Mismatches 1015; Indels 75; Gaps 7;

```

```

QY 396 TTTTCTTAATCCCGACAGTACGTACTATGATGGAGGCTATCAATGTTAAAG 455
DB 3528 TTTATGTTAGATTCTTCTTACATACGCTACGTACGAGGCGCTATGATATGAAG 3587
QY 456 GAATTTAAGCTATTGTGATGATGATGATGATGATGATGATGATGATGATGATG 515
DB 3588 CACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3647
QY 516 TGGCGGAGCAAT-----AATGTCGAGGCTCTTGAATGCAACAA 560
DB 3648 TGGCGGAGCAATGTCGACACCAAGCGAGTGGATCTCTGAATGATATCTCTACAA 3707
QY 561 TAAGTCCCTTTCTTTATGGAATAGTCTTCAACAGCTGGCGGAGCTTATACCA 620
DB 3708 TAGACTGTGATTTGCTTCAACAGCTGAGGAGGAGGCGGAGCTTATACCA 3767
QY 621 AATCTCACATATCTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 680
DB 3768 AATCTCACATATCTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3827
QY 681 TGGTGTAAAGGAGTGTATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 740
DB 3828 AATCTCACATATCTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3887
QY 741 CAGTGGCAGCATATTTTGAAGGCAATAGATGAGTACAGAGACCGTCTCT----- 795
DB 3888 GAGAGAAACATTTACTTTGTAAGAAATACCTTACAAACCGGAGATGACGATCTCC 3947
QY 796 ----CATAGTCTATTGATTTAGAACTAGCGCTAGATTAATGCTTACGTCGTCGCA 851
DB 3948 TAACTCTAATGCGATCAATAGAGAGTAAAGGAAATTCACGGAATTTACGGGCTCTAA 4007
QY 852 AGACATAGCATATCTTTATGATCCGATTAATGATGATGATGATGATGATGATGATGAT 911
DB 4008 AATATCTAATATTTCTTATGATCCATCTTCAAGGAACCTCAT-----CAAA 4061
QY 912 TGGCTCAATATTTATAGCCCTGATCTAGTATGATGATGATGATGATGATGATGATGAT 971
DB 4062 CGATTTGAGATATTAATAGCGCTCTGCGGAGCTCTCATTCATATCAAGAGATTTCT 4121
QY 972 CTTTCTGAGAGAGCTACGAGGAGAGTAAAGTATGAGAGAGACCGGACCTTTCTAA 1031

```

```

DB 4122 ATTTCTGGAGAAACCTTAACAGACAGATGAATCTTAAAGTGTGACAAATTTAAATCTTC 4181
QY 1032 ATTACTTCAAAATGTTGCTTTTAAATGGAAGTATGATTTTAAAGGAGTGTCTTT 1091
DB 4182 ATTACAGCAGCCAGTCTCTTATCCGAGAGAAATGATTTGCTTCAAAAGGAGTCTCTTT 4241
QY 1092 AAGTGGAGAGGTTTCTTCAAGATGCAAACTTAAAGTATGATTTTAAAGGAGTGTCTTT 1151
DB 4242 AGAGACACAGAGCTTCTCAAGAGGCGGTTCTCTCGGAGTGTGATTCAGAGAGCAG 4301
QY 1152 GTTGTGCAAAACCGGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 1211
DB 4302 ATTATCACTATACGCTGCGGAGTATTCATATCAAGACCTAAGATTAATGATGATGATG 4361
QY 1212 CAGGAGCGGAAAGATTAATAACTGATGTCGACAGCTCAGAAAGATTTGATGATGATG 1271
DB 4362 AGTCTTTAAGCAGCCCGCTCAGCTTAAACAGAAAGTGTCTTCAAAATTAAGTATGATG 4421
QY 1272 TGGTCTGTTGTACTGCAATTAAGCAGTATGATGATGATGATGATGATGATGATGATG 1331
DB 4422 TGGGAGGCTCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4481
QY 1332 GGAACATTCCT-----ATGATGAGATTTGATGATGATGATGATGATGATGATGATG 1370
DB 4482 TGACCAAGCTCTCTCTCTATTAATAATCAAGTGTGATGATGATGATGATGATGATGATG 4541
QY 1371 GAAGACATGCTGATTTCTGCAATTCGCAATTCGCAATTCGCAATTCGCAATTCGCAATTC 1430
DB 4542 CATCAGACGCTTATCTCTGTCCTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4601
QY 1431 CTAT-----CAGGGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1487
DB 4602 ACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4661
QY 1488 TTGGCAAAAGCAATTTTATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1547
DB 4662 TTGGCAAAAGCAATTTTATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4721
QY 1548 TCTTTGGGTTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1607
DB 4722 CTTATGGGAGTCTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4781
QY 1608 AGTGTCTCTTACGAAAGAGATTTTGGTGCAGCAGTATTCATATGTTTGCATAGAG 1667
DB 4782 TGGTATGAGAAACAAAGAGTGTGCGGTTCTCTCATGAGCAACTCTCTGATGATGAG 4841
QY 1668 CGTCTGTAATCAAGAAATTCCTGATGATGATGATGATGATGATGATGATGATGATGATG 1727
DB 4842 TGGAGATGAAAGAGTGTGCGGTTCTCTCATGAGCAACTCTCTGATGATGAG 4901
QY 1728 CAGGAGATGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1787
DB 4902 TGGTCACTCTCTTAAAGAGACCTATTTACCTTTGCGTTCTGCAATCTCTTTGCTGAGA 4961
QY 1788 CAAAGACTATTTATGATTAATCAATTTGCAAAAGACCTAAGCAGATCTTTTACGTTTGA 1847
DB 4962 CAAAGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5021
QY 1848 GCAAGATGCTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1907
DB 5022 GCACCTCTACCTCTCAACCCCAAAACTATTTGAGATTTAGAGAGAGCAAAAGTTTCTCA 5081
QY 1908 GATCTCTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1967
DB 5082 ATTCAGTATGAGAAATTTCTTATGAGAAATTTCTTATGAGAAATTTCTTATGAGAA 5141
QY 1968 CGGCACTACGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2027
DB 5142 CAGCATTTCAGCAACCGATGAGAAACGACATATACCTCATTTGCCAAGATC----- 5192
QY 2028 GACGATCATCTCTTGGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2087
DB 5193 -----CGAAGTCTTGTGAGAGCAAGATGATGATGATGATGATGATGATGATGATGATG 5246

```

```

Db 2262 CAAAGATGTTTATCGCTCAACAACTAGAACCTACGGTGAACCTTATCTTCAA 2321
QY 1848 GCAGATGCTTCCCTATACCTGTGTAGATCCCTTTAGAGAGGAGACCTCCGCGA 1907
Db 2322 GCACCTCATACCTACCAACCCCAAACTATTTGATAGGAAGCAAGGTTTCTGA 2381
QY 1908 GATCCGTGTGCTTATGTTTCCAGACTCTGCCGTCCTTCTATGGCAGCTAGCTA 1967
Db 2382 ATCAGCTATACAAAATTCCTTAGGAAATTCCTTACGCTTGATGCTCAAGTTTCT 2441
QY 1968 CGGCATACGAGATCATCGCATGAAGACCGAGCTCTTACCCCCCCCCCGACGCTCTC 2027
Db 2442 CAGCATTCAGACAACCTATGGAAGCAGCTATATCCTCATTTGCCAGAAATC----- 2492
QY 2028 GACGATCATCTCTTGTGGGAGATATGCTGGCTGGAGAGCTGGGAACTCCAGTTGC 2087
Db 2493 -----CGAAGTCTCTGGACCAAGAGTGTATGCTGTGCTGAGACCTTCC 2546
QY 2088 TGTGAAATATACAGCGGCGAGAGATTTTCCGAGATACCTCATTTGTAAGTCCA 2147
Db 2547 TTTTGTCTTCCACACCACTCTCTTTCAGACCTTCAATCCACAGATGAAGTCTGA 2606
QY 2148 AGCTGTACTCGCGCCAGATAGCTTTGTAAGTACAGCTATCACTCGTATTTAG 2207
Db 2607 AATGTTTATGATACCAAAATAGCTTCTGAAAGCTCTAGTATGCGGTGTTAG 2666
QY 2208 TGATTCGATTTTATTAACCTTGGATTCCTTGTGAATCAAGTTAGAAACGTTTGC 2267
Db 2667 TATTGGAAGGCTGTTACCTCGATTCGTGTGGTCCGAATTCGTGACGGGAGAT 2726
QY 2268 AGACCAATAT-----TATCATGTTGTAGCGATGATATCTCCAGATGTTTCTAGTAA 2321
Db 2727 CGGAGATTCCTACACCTATGATCTCTCAGATTCCTTCTCCATGCTCATTCGAACAA 2786
QY 2322 CCCCAATGATGACACTCTTCTTCCCAACCAAGGAGTGGAAAGCAAAAGTTTCGAA 2381
Db 2787 TCCCAATCTACACGACTCTTGTGATGAGCCACGACTCTGAAAAATTCGCGGTGCAA 2846
QY 2382 CTTAGCAAGAGCTGTATGTTTCAGAGCTCAGCTTTGATCTTTGGAGCTGCAGC 2441
Db 2847 TCTTTCAAGACGAGCTTTTACTGAGGGGTAGCAACAACGATCTTCAAACTCCAAATG 2906
QY 2442 AGACCTTTGCGGAACCTTGGCTTTGATGCGGGGATCTTCTGATCTTAATGTAGA 2501
Db 2907 TGAGCTCTTGGCAATATACGCTATGGAACCTCCGTGATCTTCAAGGAACATGATAGA 2966
QY 2502 TCGGGGTAGCAAAATCAAAATTTTAG 2526
Db 2967 TGTGTGTACCAAACTCGATCTCTAG 2991

```

```

FEATURES
    MEDLINE
        JOURNAL
            outer membrane proteins of Chlamydia pneumoniae
            Infect. Immun. 67 (1), 375-383 (1999)
        99081766
    SOURCE
        Location/Qualifiers
            1..6030
                /organism="Chlamydia pneumoniae"
                /isolate="CDC/CWL-029/VR-1310"
                /db_xref="taxon:83558"
            /clone="pEX2-18"
            /clone="B8F3"
            /clone_id="pbluescript"
            /clone_id="pEX2"
            /dev_stage="elementary body"
            complement(1..2804)
            /gene="omp5"
            complement(<1..2742)
            /gene="omp5"
            /note="putative lipoprotein"
            /codon_start=1
            /transl_table=11
            /product="outer membrane protein 5"
            /protein_id="CAA04671.1"
            /db_xref="GI:325935"
            /db_xref="SPTREMBL:086163"
            /translation="MKSQFSLVLSYLAFCSTVFPAATAENIGPSDFSDSTNTG
            TYPKNTTGTGIDTLTGIDTLONLGDALTKGCFSDPTESLSFAGYSLSPNLS
            SAEBAALSVTIDKRLSLTGPSILFLAAPSIVTPSGKAVCGGDLTFDNNGTILF
            KQDYCENGAISTKSLKNSGSISEFNKSSATKKKGAICATCTVITNTATP
            LFSNLNLEAGGAINSTGCTTGMSLVSESNVITATNGAGLSADAVTISGNS
            VTSGNVAANGAIVAKKLTLLASGGGSISSNVIQGTAGAGLSIISLAEGSL
            SAEGDITTFNGMAIVATPOTTRKNSIDISTAKITNLRAISGHSIFPYDPIANTAA
            DSTDTLNLINKADGNSDYSISYFSGEKSDEBAKYADNLSTLKPVTYLTGNLVL
            KRGTLYDTKGTGTAGSSVIMDAAGTLKASTEEVTLGLSIPVDSLEGKRYIAASA
            ASKNVALSGPILILLDNGNAVENHMDGTLKQDFSVOLSACTGATTTDDVPVYAAPPT
            HIGVQGTGWTWDDTASTPKRTATLANTNTGTLRPERQGPLVPNSLSGSSDIOA
            IQGVIERSAULTLSDRGFWAGVANEFDKDKGKKRYRRAKSGVYALGGAQTCSENL
            ISFAFCOLFGSBDKFLVAKNHTDPTAGAFYIOHTTECSGFIQLDLKLPESMSKSEPL
            LECOLAYSHVSNLDTKTYTAYPEVKGSGMNNANPMIMIGASSHSYPEYLHCFDYAPAYI
            KLNLTJYIRDSFSEKGTGERSFDSNLFNLSTLPGLGVFEKFSQNDSPSYLTLSYVD
            LIRNDPKCTALVYISGASMETYANNLRLQYRAGSHYAFSPMFEVLQGFVEYVRS
            "
            complement(1..2697)
            /gene="omp5"
            /product="outer membrane protein 5"
            complement(2698..2742)
            /gene="omp5"
            complement(2747..2752)
            /gene="omp5"
            /note="putative"
            complement(2777..2782)
            /gene="omp5"
            /note="putative"
            complement(2799..2804)
            /gene="omp5"
            2863..5726
            /gene="omp4"
            2863..2868
            /gene="omp4"
            /note="putative"
            2896..2902
            /gene="omp4"
            2905..2955
            /note="putative"
            /gene="omp4"
            2905..5681
            /gene="omp4"
            /note="putative lipoprotein"
            /codon_start=1
            /transl_table=11
            /product="outer membrane protein 4"
            /protein_id="CAA04672.1"
            /db_xref="GI:325936"
            /db_xref="SPTREMBL:086164"

```

```

RESULT 7
LOCUS AB1827 3200 bp DNA PAT 21-JAN-2000
DEFINITION Sequence 1 from Patent WO9858953.
ACCESSION AB1827
VERSION AB1827.1 GI:6731863
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3200)
AUTHORS Madsen, A. and Birke Lund, S.
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL Patent: WO 9858953-A 1 30-DEC-1998;
MADSEN ANNA SØFIE (DK); BIRKE LUND SVEND (DK)
FEATURES
    source
        1..3200
            /organism="unidentified"
            /db_xref="taxon:32644"
            205..>2987
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAB69199.1"
            /translation="MKTSPVWLVSVLAFSCHQSLANELLSPPDSFNCNIDSGT
            TPSTATPYSLTGVDFEPKGTPLSDSCFKOTDNLTFNGHSLTFEGIDAGTHA
            GAATSTANKMLTFSGPFLSPDSSPTVTTCGGTCTTACAGVNLNTRKTVVGNFS
            TADGAIKGFLLTGTSGDALFSSNSSSTGGAIAATAGARIANNCTVPLSNIAS
            TSGAIDDEGTSLISNNKFLYFEGMAATVGAICNTKASPELILSNNTLTPASN
            VAETSGAIIHAKKLALSFGTEFLRNWVSATRGKAIISDASGEISLDETQITP
            VRNTLTGTGDTFKRNAINIGSKTEFLPAAAHNTTFEYDPTSEBTSDELKINN
            GSAGLNVGCTILFSGETLPADBLKVADNKSEFTQPSVSGKLLQKQTELEST
            FSOEGSLILGMDGTLSTTAGSITITMLGINVDSIGKOPVSLAKASAKRVISGK
            LNLIDRSNIYESHESHDOLFSLIKITVDVDPVNDISLIYPAEDPSPSEFQO
            QWVNTWTDPTATKREARTATKTKGFPSPERKSAIVCNTLMGFTDRLSLQVETG
            ATGHEHKGQEVWSSMTNLFHKTGDNKRGFRHTSGGIVTGSATPKDLETFACHL
            FARKDEFIANNNSRTNYGGTLFPHSHSLQPNLRLGRASFSALEKFEPLPLAL
            DVOYFSHSDNRMETHYSLPESBSWNECIAGIGIDLPFVLSNPILKRTIPOM
            KVENWYSQNSFESSSDGRGFSIGRLNLEIPYGAKEVQGDIGSDSYDLSGEFVS
            YVRNPOSTATLVMSPDMSWKIRGGLNSQAFRLNGSNMYVNSNCELEGHYAMELRGS
            SRNTNVYGTFLR"
BASE COUNT 946 a 708 c 623 g 923 t
ORIGIN
Query Match 13.5% Score 341; DB 9; Length 3200;
Best Local Similarity 50.1%; Pred. No. 6.3e-72;
Matches 1095; Conservative 0; Mismatches 1015; Indels 75; Gaps 7;
396 TTTCTTAATCCCGACAGTACGATGATGAGGAGGAGCTATCATGTTAAAG 455
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
828 TTAATGATTCCTATCAACATAGCGCTACGACAGAGCGCATGATGATGAAG 887
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
456 GAATTTAAAGCTATTGATATGATGATGATGATGATGATGATGATGATGAT 515
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
888 CACGTCGATACATCAACAAATTTCTATATTTTGAAGGAATGACGCAAAACTAC 947
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
516 TGGCGGAGCAAT-----AATGTGACAGGCTCTTGAAGATGGCAACAA 560
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
948 TGGCGGTGCGATCTCAACACCAAGCGAGTGATCTCTGATGATGATATCTTACAA 1007
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
561 TAAGTCCCTTTCTTTATTTGAATAGTTCTTCAACAGCTGGCGGAGGATTCATACCA 620
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1008 TAAAGCTGATCTTTGCTTCAAAAGCTAGCAGAAACAAAGCGGTGGCGCATTCATAGCTAA 1067
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
621 AAACCTCACATCTCTTCTGGTGGGAAACTATTTTCAAGGGGAATACAGCCCTACAGGC 680
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1068 AAAGGTAGCCCTTCTCTGGAGGCTTTACAGAGTTTACAAATAATAGTCTCATACAC 1127
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
681 TCGTGTAAAGAGGAGTGCATGCGATGACAGATCTGTGACACCTATCATCTTTGAGAGA 740
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1128 AACTCTTAAGGGGGGTGCTATAGCATGATGCTCAGAGAGGCTCAGTCTTTCTGACAGA 1187
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

741 CAGTGGCGACATTAATCTTTGAAGCAATACGATAGAGCTACAGAACCTCTCT----- 795
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1188 GACAGGAAACATTAATCTTTGTAAGAAATACCTTACAAACACCGGAGATACGATATCC 1247
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
796 ---CATAGTGTATTTGATTTAGAACTAGCCCTAAGATTAATCTGCTTACGCTGCGCA 851
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1248 TAAAGGTATGCAATACATATGAGAAAGTAAACGGGAAATCAGGAAATTCGGGCTGTCTAA 1307
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
852 AGGACATCGATATCTTTATGATCCGATTTCTGTAACAGGATGACATCTGTTGCTGA 911
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1308 AATCATACAAATTTCTTCTATGATCCATCTTCACAGAGAACCTCTAT-----CAGA 1361
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
912 TCGTCTCAATATTAATAGCCCTGATACTGAGATACAAAGAGATATACGGAACATACT 971
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1362 CCTATTGAGATTAATTAACGGCTCTGCGGAGACTCTCATATCATATCAAGACATCT 1421
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
972 CTTTCTGAGAGAAAGCTCACGAGGACAGACCTAAGATGAGAGAACCGACCTCTTAA 1031
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1422 ATTCTTGAGAGAAACCTTAACAGAGATGACTTAAAGTTGCTGACAAATTTAAATCTTC 1481
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1032 ATTACTTCAAAATGTTGCTTTTAAAAATGAGACTGATGTTTAAAAAGTATGCTGTTT 1091
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1482 ATTCACGACGACGCTCTCCCTATCCGAGAGAAAGTATTTGCTACAAAAGGAGTACCTTT 1541
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1092 AAGTCGAAAGGCTTCTCTCAGAGATGCAAACTCTAAGTTGATTAATGATTAAGGACGTC 1151
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1542 AGAGAGACAGACCTCTCTCAAGAGCGCGCTCTCTCCCGGACATGATTAACGAGACGAC 1601
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1152 GTTGTGTCAAACACCGAAAGTATGAGTTAAAGAAATTTGAAATTAATATAGACTCTCT 1211
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1602 ATTATCACTACAGCTGAGGAGTATTAACAATCAGAACCTAGGAAATCAATGATGACTCTCT 1661
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1212 CAGGAACGGGAAAGATTAATCAATCAGTCCGACAGCTGAGAAAGATATTCGATAGA 1271
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1662 AGCTCTTAAGCAGCCCGCTACGCTTAACGAAAGAGTCTTTAAATTAAGTATGCTATAC 1721
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1272 TCGTCTGTTGACTGGCAATTAAGCATGAGAGTTTATTCAAAATGCGCTTTTGAATGA 1331
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1722 TGGGAAGCTCAACGATGATGATTAAGGAAGCAATTTATGAAGTCAATATGTTACGCA 1781
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1332 GGACCATTCCT-----ATGATGGGATTCCTTGATGATGATGCTG 1370
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1782 TGACCAAGCTCTCTCTCTATTAATAATCAAGGTTGATGCTGATGATGATCACTAACGTTGA 1841
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1371 GAAAGCATCGATGATTTCTGCAATTCGACAGTATTAATGCTGATACATCTCCGATGAG 1430
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1842 CATGACGACCTTATTCCTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1901
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1431 CTAT---CAGGAAAGTGGACAAATCAATGATGATGATGATGATGATGATGATGATGATGAT 1487
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1902 ACAATGGAATGTTAATTTGACTACAGGATACAGCTACAAATACAAAAGAGCCACGGCAAC 1961
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1488 TTGGCAAAAGCAATTTTAATCCACATGCTGACAGAGGCTCCGTTAGTTCTTAATCT 1547
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1962 TTGGACCAAAACAGAGATTTGTTCCAGCCCGCAAAAGAAATGTGGTTAGTATGCAATAC 2021
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1548 TCTTTGGGCTCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1607
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2022 CCTATGGGAGAGCTTTACGACATTCGCTCTGCAACAGCTTGTAGAGATGGCGGCAAC 2081
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1608 AGTGCTCTCTTACGAAAAGAGATTTGGGTTGACAGCAATTTCCAAATGTTTTCATAGAG 1667
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2082 TGGTATGGAACCAAAAGATTTCTGGGTTTCTTCATGAGCAAGCAATCTCTCATATAGAC 2141
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1668 CGGTGCTGTAATTAAGGAATTTCCGTCATGATGATGATGATGATGATGATGATGATGATGAT 1727
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2142 TGGAGATGAAATTCGAAAGGCTCTCGTATACCTCTGAGGCTACGCTATGAGTGGAGAG 2201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1728 CACGAGATATCGCGGTGATACCTTCTCTGAGGTTTTCGCTACGCTCTTTCGCGGTA 1787
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2202 TGTCTACACTCTTAAGACACCTATTTACCTTTCGCTTCGCAATCTCTTGTGATAGANA 2261
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1788 CAAGAATCTATTATGAAATACCAATTTGCAAAAGACCTACGAGAGATCTTTACGTTTGA 1847
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

2521 TTTTAG 2526  
|||||  
Db 20782 TTTAG 20787

RESULT 6  
LOCUS A81855 1830 bp DNA PAT 21-JAN-2000  
DEFINITION Sequence 29 from Patent WO9858953.  
ACCESSION A81855  
VERSION A81855.1 GI:6731880  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 1830)  
AUTHORS Madsen, A. and Birkelund, S.  
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE  
JOURNAL Patent: WO 9858953-A 29 30-DEC-1998;  
MADSEN ANNA SORIE (DK); BIRKELUND SVEND (DK)  
FEATURES  
source Location/Qualifiers  
1..1830  
/organism="unidentified"  
/db\_xref="taxon:32644"  
<1..>1830  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAB69202.1"  
/db\_xref="GI:6731881"  
/translation="DLTIGSRDSYNGDTSTTEFPKRAVSDASGTYIILDGVYSISOA  
GRQSLTSCSFNTAGNLTFLNGFSLHFDNIISSTVAGVAVDNLNTSPDTGDKREY  
LRMLAPRTTGKAIKITDGLVPESTIGNDPIVTGSTSVADNLNTSPDTGDKREY  
GTVSEGERLTEREAKDEKRNKSLKLNQAFKNGTYVLKGDVLSANGSODNSLI  
MDLGLVANTESIELNLEINIDSLRNKIKLSAATOKDIRIPVLAISDSF  
YONGFLNEHSDYDGLLELDAGKDVISADSRSIDAVQSPYQKMTIMSDDKAT  
VSWAKQSPNPTAEQEAFLVPLNLMGSFIDVRSPONFELGTBGAPEKRFVAGISN  
LHRSRENORKEFHVSGAVVGAFTMGADTSLGFAOLFARDKDYFMNTNFKTYA  
GSLRLQHDASLSVSVIILGEGGLEILPYVSNLPCSPYQGLSYGHDHMKTESL  
PPRPRLSDPHTSMGCVYVAGELCTPVAVENSGRFRFRFPYVVOAVYSRQDSFV  
ELGAIKRDSDSLVNLALPILGKIKLEKRAEQTHVVAHMSPDV"

BASE COUNT 485 a 357 c 443 g 545 t  
ORIGIN

Query Match 56.4%; Score 1425.2; DB 9; Length 1830;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1430; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

873 TGATCGATTACTGTACAGAGATGACATCTGTGTGATGCTCTCAATATTAATAGCCC 932  
|||||  
Db 393 TGATCGATTACTGTACAGATGACATCTGTGTGATGCTCTCAATATTAATAGCCC 452

933 TGATCTGAGATPAACAAGATATACGGGAACCAATAGCTTTCTTGAGAGAAGCTCAC 992  
|||||  
Db 453 TGATCTGAGATPAACAAGATATACGGGAACCAATAGCTTTCTTGAGAGAAGCTCAC 512

993 GGGAGGAGAGCTAAAGATGAGAAGAACCCACTTCTAAATTAATTAATGTTGCTTT 1052  
|||||  
Db 513 GGGAGGAGAGCTAAAGATGAGAAGAACCCACTTCTAAATTAATTAATGTTGCTTT 572

1053 TAAAAATGGAGCTGTATGTTTAAAGGTGATGCTTTTAAAGTGCAGAACGTTTCTCTCA 1112  
|||||  
Db 573 TAAAAATGGAGCTGTATGTTTAAAGGTGATGCTTTTAAAGTGCAGAACGTTTCTCTCA 632

1113 GGATGCAAACTCTAAGTTGATATGATTTAGGAGCTGCTGGTTGCAAAACCGAAG 1172  
|||||  
Db 633 GGATGCAAACTCTAAGTTGATATGATTTAGGAGCTGCTGGTTGCAAAACCGAAG 692

1173 TATCGAGTTTACGAATTTTGAATTAATATAGACTCTCTCAGAGAGCGGAAAAAGATAAA 1232  
|||||  
Db 693 TATCGAGTTTACGAATTTTGAATTAATATATAGACTCTCTCAGAGAGCGGAAAAAGATAAA 752

1233 ACTCAGTCTGCTCCACAGCTCAGAAAGATATTCGTATAGATGCTCTGTACTGCGCAAT 1292  
|||||  
Db 753 ACTCAGTCTGCTCCACAGCTCAGAAAGATATTCGTATAGATGCTCTGTACTGCGCAAT 812

1293 TAGCGATGAGATTTTATCAAAAATGCTTTTGAATGAGAGACCATTCCTATGATGGAT 1352  
|||||  
Db 813 TAGCGATGAGATTTTATCAAAAATGCTTTTGAATGAGAGACCATTCCTATGATGGAT 872

1353 TCTTGAGTTGATGCTGGGAAAGACATCGATTTTCGAGATTCCTGCGAGTAAATTAAT 1412  
|||||  
Db 873 TCTTGAGTTGATGCTGGGAAAGACATCGATTTTCGAGATTCCTGCGAGTAAATTAAT 932

1413 TGTAACAATCTCCGTATGCTATCAGGGAAGTGACATCAATGGTCTACTAGATATAA 1472  
|||||  
Db 933 TGTAACAATCTCCGTATGCTATCAGGGAAGTGACATCAATGGTCTACTAGATATAA 992

1473 GAAAGCTACGGTTTCTTGGGCAAGCAAAAGTTTATCCACATCTGTCAGCAGAGGCTCC 1532  
|||||  
Db 993 GAAAGCTACGGTTTCTTGGGCAAGCAGAGTTTAAATCCACATCTGTCAGCAGAGGCTCC 1052

1533 GTTACTCTCAATCTCTTGGGGTCTTTTATGATGTTGCTGCTCCAAATTTTAT 1592  
|||||  
Db 1053 GTTACTCTCAATCTCTTGGGGTCTTTTATGATGTTGCTGCTCCAAATTTTAT 1112

1593 AGAGCTAGTACTGAAGGCTCTCTTACGAAAAGAGATTTTGGGTTGACAGGCAATTTCAA 1652  
|||||  
Db 1113 AGAGCTAGTACTGAAGGCTCTCTTACGAAAAGAGATTTTGGGTTGACAGGCAATTTCAA 1172

1653 TGTATTTCATAGAGCGGCTGCTGAAAATCAAAGAAATTCGTCATGTAGTGAAGGCTGC 1712  
|||||  
Db 1173 TGTATTTCATAGAGCGGCTGCTGAAAATCAAAGAAATTCGTCATGTAGTGAAGGCTGC 1232

1713 TGTATGAGGCTATGACAGAGAGATGCGGGGTGGATACCTTGCTGCTGGGTTTGCTCA 1772  
|||||  
Db 1233 TGTATGAGGCTATGACAGAGAGATGCGGGGTGGATACCTTGCTGCTGGGTTTGCTCA 1292

1773 GCTCTTGGCCGCTGACAAAGACTTATGAATACCAATTTGCAAAAGACTACGAGG 1832  
|||||  
Db 1293 GCTCTTGGCCGCTGACAAAGACTTATGAATACCAATTTGCAAAAGACTACGAGG 1352

1833 ATCTTTACGTTTGACAGACATGCTTCCATATCTGTGTGATATCTCTTTTATAGAGA 1892  
|||||  
Db 1353 ATCTTTACGTTTGACAGACATGCTTCCATATCTGTGTGATATCTCTTTTATAGAGA 1412

1893 GGGAGGACCTCCGAGATGCTGTTGCTTATGTTTCCAAAGCTGCGGCTCTTTCTA 1952  
|||||  
Db 1413 GGGAGGACCTCCGAGATGCTGTTGCTTATGTTTCCAAAGCTGCGGCTCTTTCTA 1472

1953 TGGGCACTTAGCTACAGGCAATACGATATCGCATGAAGACCGAGTCTCTACCCCCC 2012  
|||||  
Db 1473 TGGGCACTTAGCTACAGGCAATACGATATCGCATGAAGACCGAGTCTCTACCCCCC 1532

2013 CCCCCGAGCTCTCGACGATCATACTTCTTGGGAGGATATGTCTGGGCTGAGAGCT 2072  
|||||  
Db 1533 CCCCCGAGCTCTCGACGATCATACTTCTTGGGAGGATATGTCTGGGCTGAGAGCT 1592

2073 GGGAACTCGAGTTCGTTGAAAATATACAGGCGAGAGATTTTTCGAGAGTACCTCC 2132  
|||||  
Db 1593 GGGAACTCGAGTTCGTTGAAAATATACAGGCGAGAGATTTTTCGAGAGTACCTCC 1652

2133 ATTTGTAAAAGTCCAAAGCTTTTACTCGCCCAAGATAGCTTTGTGAATAGAGTAT 2192  
|||||  
Db 1653 ATTTGTAAAAGTCCAAAGCTTTTACTCGCCCAAGATAGCTTTGTGAATAGAGTAT 1712

2193 CAGTCTGATTTTATGATATCGATCTTTATTAACCTTGCATTCCTTGGAAATCAAGTT 2252  
|||||  
Db 1713 CAGTCTGATTTTATGATATCGATCTTTATTAACCTTGCATTCCTTGGAAATCAAGTT 1772

2253 AGAGAAAGGTTTGCAGAGATATTAATCAATGTGATGAGATATATCTCCAGATGTT 2310  
|||||  
Db 1773 AGAGAAAGGTTTGCAGAGATATTAATCAATGTGATGAGATATATCTCCAGATGTT 1830

|||||  
Db 18563 TTAGCAATATGATGATCAACCAAGGCTCTGGAGCTGATATGGAAGTGAAGAGCTAAAT 18622  
QY 361 AAGACAGTACGTTATACAGAGATTTTGGCACTTTCTTTCTTAATATCCCAAGATACA 420  
Db 18623 AAGACAGTACGTTATACAGAGATTTTGGCACTTTCTTTCTTAATATCCCAAGATACA 18682  
QY 421 GTGACATATGATGGAGCTATCAATGTTAAAGGAATTTAAAGCTATATGATATGAT 480  
Db 18683 GTGACATATGATGGAGCTATCAATGTTAAAGGAATTTAAAGCTATATGATATGAT 18742  
QY 481 AAGATATGATTCAGAGCAATTTCTCAACAGAGATGGCGAGCAATTAATTTGTCAGGC 540  
Db 18743 AAGATATGATTCAGAGCAATTTCTCAACAGAGATGGCGAGCAATTAATTTGTCAGGC 18802  
QY 541 TCCCTGAATATGGCAAAACAATAAGTCCCTTTCTTTATTTGGAATATGTTCTTCAACAGT 600  
Db 18803 TCCCTGAATATGGCAAAACAATAAGTCCCTTTCTTTATTTGGAATATGTTCTTCAACAGT 18862  
QY 601 GCGGAGCATTCATACCAAAAACTCACACTATCTTGTGGGGAACCTGATTTTCAAG 660  
Db 18863 GCGGAGCATTCATACCAAAAACTCACACTATCTTGTGGGGAACCTGATTTTCAAG 18922  
QY 661 GGGATATACAGCGCTACAGGCTCTGGTAAAGAGAGTGTATCGCATTCAGACTCTGCG 720  
Db 18923 GGGATATACAGCGCTACAGGCTCTGGTAAAGAGAGTGTATCGCATTCAGACTCTGCG 18982  
QY 721 ACCCTATCATTTCTGGAGACAGATGGCGACATTAATCTTGAAGGCAATACGATAGAGCT 780  
Db 18983 ACCCTATCATTTCTGGAGACAGATGGCGACATTAATCTTGAAGGCAATACGATAGAGCT 19042  
QY 781 ACAGAACCGTCTCATATAGTGTCTATGATTTAGAACTAGCGCTAAGATTAAGTGGCTTA 840  
Db 19043 ACAGAACCGTCTCATATAGTGTCTATGATTTAGAACTAGCGCTAAGATTAAGTGGCTTA 19102  
QY 841 CGTGTGCGCAAGAGATACGATATCTTTATGATCCGATTAAGTAAACAGATGAGCA 900  
Db 19103 CGTGTGCGCAAGAGATACGATATCTTTATGATCCGATTAAGTAAACAGATGAGCA 19162  
QY 901 TCTGTGTGATGCTCTCAATATTAATAGCCCTGATACGAGATTAACAAAGATATAG 960  
Db 19163 TCTGTGTGATGCTCTCAATATTAATAGCCCTGATACGAGATTAACAAAGATATAG 19222  
QY 961 GGAACCATATGCTTTTCTGAGAGAAAGCTACAGAGAGCAAGCTAAGATGAGAAAG 1020  
Db 19223 GGAACCATATGCTTTTCTGAGAGAAAGCTACAGAGAGCAAGCTAAGATGAGAAAG 19282  
QY 1021 CGCACTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080  
Db 19283 CGCACTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19342  
QY 1081 GATGTGCTTTAAGTGGAGAGGTTCTCTCAAGATGCAAACTCAATGATTAATGAT 1140  
Db 19343 GATGTGCTTTAAGTGGAGAGGTTCTCTCAAGATGCAAACTCAATGATTAATGAT 19402  
QY 1141 TTAGGAGACGCTGTTGGTTCAAACACCGAAGATATCGAGTAAACGAATTTGGAATTAAT 1200  
Db 19403 TTAGGAGACGCTGTTGGTTCAAACACCGAAGATATCGAGTAAACGAATTTGGAATTAAT 19462  
QY 1201 ATAGACTCTCTCAAGAGAGGGAAGAAAGATTAACATCAGTGTGCCACAGCTCAGAAAGAT 1260  
Db 19463 ATAGACTCTCTCAAGAGAGGGAAGAAAGATTAACATCAGTGTGCCACAGCTCAGAAAGAT 19522  
QY 1261 ATTGCTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
Db 19523 ATTGCTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19582  
QY 1321 TTTTGAATGAGAGACATTCCTATGATGAGATTTCTGAGTATGATGCTGGAAGACATC 1380  
Db 19583 TTTTGAATGAGAGACATTCCTATGATGAGATTTCTGAGTATGATGCTGGAAGACATC 19642  
QY 1381 GTGATTTTGCAGATTTCTGCGAGTATTAATGCTGTACAAATTCCTGATGCTATCAGGGA 1440  
|||||

Db 19643 GTGATTTTGCAGATTTCTGCAATATAGATGCTGTACAAATTCCTGATGCTATCAGGGA 19702  
QY 1441 AAGTGACATCAATTTGCTGCTATGATTAAGAAAGCTACGTTTCTTGGCAAGCA 1500  
Db 19703 AAGTGACATCAATTTGCTGCTATGATTAAGAAAGCTACGTTTCTTGGCAAGCA 19762  
QY 1501 AGTTTAAATCCACATGCTGAGAGAGAGGCTCGTGTAGTTCCTAATCTTTGGGGTCT 1560  
Db 19763 AGTTTAAATCCACATGCTGAGAGAGAGGCTCGTGTAGTTCCTAATCTTTGGGGTCT 19822  
QY 1561 TTTATATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620  
Db 19823 TTTATATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19882  
QY 1621 GAAAAGATTTTGGTGTGAGGCAATTTCAATGTTTTCATAGAGGCGGTGCTGAAT 1680  
Db 19883 GAAAAGATTTTGGTGTGAGGCAATTTCAATGTTTTCATAGAGGCGGTGCTGAAT 19942  
QY 1681 CAAAGGAATTCGCTCATATGATGAGGAGTGTGATGATGCTTACAGAGATGCGG 1740  
Db 19943 CAAAGGAATTCGCTCATATGATGAGGAGTGTGATGATGCTTACAGAGATGCGG 20002  
QY 1741 GGTGTGATATCTTGTCTGTGGGTTTGTCTCAAGCTTTTCCGCTGACAAAGACTTT 1800  
Db 20003 GGTGTGATATCTTGTCTGTGGGTTTGTCTCAAGCTTTTCCGCTGACAAAGACTTT 20062  
QY 1801 ATGAATACCAATTTGCAAAAGCTACGAGATCTTACGTTTGGAGACGATGCTTCC 1860  
Db 20063 ATGAATACCAATTTGCAAAAGCTACGAGATCTTACGTTTGGAGACGATGCTTCC 20122  
QY 1861 CTATACCTGTGTGATATCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
Db 20123 CTATACCTGTGTGATATCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 20182  
QY 1921 TATGTTTCCAAAGACTGTGCGGCTCTTTCTATGAGGCACTTAAGTACGAGAT 1980  
Db 20183 TATGTTTCCAAAGACTGTGCGGCTCTTTCTATGAGGCACTTAAGTACGAGAT 20242  
QY 1981 CATGCGATGAAGACGAGCTGTACACCCCGCCCGAGCGTCTGAGAGGATCATTA 2040  
Db 20243 CATGCGATGAAGACGAGCTGTCTA -CCCCCCCCCCCCGAGCTCTGAGAGGATCATTA 20301  
QY 2041 TCTTGGGAGATATGCTGTGGCTGAGAGCTGGAACCTGGAATCTTTGAAATAC 2100  
Db 20302 TCTTGGGAGATATGCTGTGGCTGAGAGCTGGAACCTGGAATCTTTGAAATAC 20361  
QY 2101 ACGGAGAGAGATTTTCCGAGAGTACACTTATTTAAAGTCCAACTGTTACTCG 2160  
Db 20362 ACGGAGAGAGATTTTCCGAGAGTACACTTATTTAAAGTCCAACTGTTACTCG 20421  
QY 2161 CGCAGAGATAGCTTGTGATAGAGGCTATCAGTCGATTTTGTGATTCGACTTT 2220  
Db 20422 CGCAGAGATAGCTTGTGATAGAGGCTATCAGTCGATTTTGTGATTCGACTTT 20481  
QY 2221 TATTAACCTTGCATTCCTTGTGAATCAAGTTAGAGAAAGGTTTGCAGAGCAATTAAT 2280  
Db 20482 TATTAACCTTGCATTCCTTGTGAATCAAGTTAGAGAAAGGTTTGCAGAGCAATTAAT 20541  
QY 2281 CATGTTGTAGCATATATCTCAAGATTTTGTCTGATAGTAAACCCCAATATGAGACTAC 2340  
Db 20542 CATGTTGTAGCATATATCTCAAGATTTTGTCTGATAGTAAACCCCAATATGAGACTAC 20601  
QY 2341 CTACTTTCCAAACGAGGAGTTGGAAGACCAAGGTTGCAACTTACAGAGAGAGGCTGT 2400  
Db 20602 CTACTTTCCAAACGAGGAGTTGGAAGACCAAGGTTGCAACTTACAGAGAGAGGCTGT 20661  
QY 2401 ATTGTTCAAGGCTCAGGTTTTCATCTTGGAGCTGACAGAGGCTTTTCGGAACTTT 2460  
Db 20662 ATTGTTCAAGGCTCAGGTTTTCATCTTGGAGCTGACAGAGGCTTTTCGGAACTTT 20721  
QY 2461 GCGTTTGAATGCGGAGGATCTTCTGTACTATTAATGATGAGCGGATGCAAAATCAAA 2520  
Db 20722 GCGTTTGAATGCGGAGGATCTTCTGTACTATTAATGATGAGCGGATGCAAAATCAAA 20781  
|||||





```

Db 10298 GAAAAAGATTGGTGGTCAGGACATTTCCAAATGTTTGCATAGAGAGCGTGCTGTAAT 10239
Qy 1681 CAAGAANAATTCGTCATGATGATGAGTGGTCTGTAGTGGCTGTGACAGAGATGCCG 1740
Db 10238 CAAAGGAAATTCGTCATGATGATGAGTGGTCTGTAGTGGCTGTGACAGAGATGCCG 10179
Qy 1741 GGTGGTGAATACCTGTGTCTGGGTTTGGTCACTCTTGGCGGTGACAAAGACTACTT 1800
Db 10178 GGTGGTGAATACCTGTGTCTGGGTTTGGTCACTCTTGGCGGTGACAAAGACTACTT 10119
Qy 1801 ATGAATATCAATTTCCGCAAGACCTACGAGATCTTTACGTTTGCAGACAGATGCTTCC 1860
Db 10118 ATGAATATCAATTTCCGCAAGACCTACGAGATCTTTACGTTTGCAGACAGATGCTTCC 10059
Qy 1861 CATATCTCTGTGTGATGATCTCTTTTAGAGAGAGGAGACTCCGCGAGATCTCTGTGCT 1920
Db 10058 CATATCTCTGTGTGATGATCTCTTTTAGAGAGAGGAGACTCCGCGAGATCTCTGTGCT 9999
Qy 1921 TATGTTTCCAAAGACTGTGCGCTCTTCTATGAGGACACTTACGAGCCATACGAT 1980
Db 9998 TATGTTTCCAAAGACTGTGCGCTCTTCTATGAGGACACTTACGAGCCATACGAT 9939
Qy 1981 CATGCGATGAAGACGAGTCTCTACCCGCCGCCGCCGACGCTCTGACGAGATCACTACT 2040
Db 9938 CATGCGATGAAGACGAGTCTCTACCCGCCGCCGCCGACGCTCTGACGAGATCACTACT 9879
Qy 2041 TCTTGGGGAGAGATATGTCGGGCTGGAGAGCTGGGACACTCGATGCTGTGTAATATAC 2100
Db 9878 TCTTGGGGAGAGATATGTCGGGCTGGAGAGCTGGGACACTCGATGCTGTGTAATATAC 9819
Qy 2101 AGCGGAGAGATTTTCCGAGAGTACACTCCATTTGTAAGATCCAAAGCTGTTACTCG 2160
Db 9818 AGCGGAGAGATTTTCCGAGAGTACACTCCATTTGTAAGATCCAAAGCTGTTACTCG 9759
Qy 2161 CGCGAAGATAGCTTTGTATAGTAGAGTATCAGTCTGATTTTATGATTCGACTCTT 2220
Db 9758 CGCGAAGATAGCTTTGTATAGTAGAGTATCAGTCTGATTTTATGATTCGACTCTT 9699
Qy 2221 TATTAACCTTGGCGATCTCTGTAATCAAGTTAGAGAAAGGTTTGCAGAGCAATTTAT 2280
Db 9698 TATTAACCTTGGCGATCTCTGTAATCAAGTTAGAGAAAGGTTTGCAGAGCAATTTAT 9639
Qy 2281 CATGTTGATGAGCATGATTTCTCCAGATGTTTGTCTGATAGTACCCCAATGATGACTACC 2340
Db 9638 CATGTTGATGAGCATGATTTCTCCAGATGTTTGTCTGATAGTACCCCAATGATGACTACC 9579
Qy 2341 CTACTTTCACACCAAGGAGGATTGGAAGACCAAGGTTTGCAGACTTACCAAGACAGCTGCT 2400
Db 9578 CTACTTTCACACCAAGGAGGATTGGAAGACCAAGGTTTGCAGACTTACCAAGACAGCTGCT 9519
Qy 2401 ATTTGTTCAAGGCGTCAAGTTTGGATCTTTGGAGCTGCGAGAGCTTTTGGGAACTTT 2460
Db 9518 ATTTGTTCAAGGCGTCAAGTTTGGATCTTTGGAGCTGCGAGAGCTTTTGGGAACTTT 9459
Qy 2461 GGGTTTGAATGGCGGAGATCTTCTCTAGCTATATATGATGATGCGGGTACCAAAATCAAA 2520
Db 9458 GGGTTTGAATGGCGGAGATCTTCTCTAGCTATATATGATGATGCGGGTACCAAAATCAAA 9399
Qy 2521 TTTTATG 2526
Db 9398 TTTTATG 9393

```

```

RESULT 5
AP002545 299650 bp DNA BCT 08-DEC-2000
LOCUS Chlamydia pneumoniae J138 genomic DNA, complete sequence,
DEFINITION section 1/4.
ACCESSION AP002545 AB033780 AB033781 AB033792 AB033793 AB033794 AB033795
AB033796 AB033797 AB033798 AB033799 AB036071 AB036072 AB036073
AB036074 AB036075 AB036076 AB036077 AB036078 AB036079 AB038346
AB038347 BA000008

```

```

VERSION AP002545.2 GI:9956082
KEYWORDS Chlamydia pneumoniae J138 (strain:J138) DNA.
SOURCE Chlamydia pneumoniae J138
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (sites)
AUTHORS Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
Shiba,T., Ishi,K., Hattori,M., Kuhara,S., and Nakazawa,T.
TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CRL029 from USA
JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)
MEDLINE 20330349
REFERENCE 2 (bases 1 to 299650)
AUTHORS Shirai,M.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.
Mutsunori Shirai, Yamaguchi University School of Medicine,
Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi
755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp,
Tel:81-836-22-2227, Fax:81-836-22-2415)
COMMENT On Aug 31, 2000 this sequence version replaced gi:6172286
gi:6172288 gi:6172310 gi:6172312 gi:6172314 gi:6172316 gi:6172318
gi:6172320 gi:6172322 gi:6172324 gi:6635158 gi:6635160 gi:6635162
gi:6635164 gi:6635166 gi:6635168 gi:6635170 gi:6635172 gi:8547422
gi:8547426 gi:8547430 gi:8978373.
AB033780-AB033781, AB033792-AB033799: Submitted (25-Oct-1999)
AB038345-AB038347: Submitted (14-Feb-2000)
AB036071-AB036078: Submitted (18-Dec-2000).
FEATURES
source
1..299650
/organism="Chlamydia pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/note="synonym:Chlamydia pneumoniae (strain J138)"
complement(1..282)
/gene="CPJ0001"
complement(1..282)
/gene="CPJ0001"
/codon_start=1
/transl_table=11
/product="Ct001 hypothetical protein"
/protein_id="BA98211.1"
/db_xref="GI:8978374"
/translation="MGKIRGLSLIYLICALNVLGIITHNKLIITAKLGVSVP
AQLYIILIGIAGVYILSLFPCPSKRSRSHSDSCSGCHSHSDKN"
573..878
/gene="gac"
573..878
/gene="gac"
573..878
/codon_start=1
/transl_table=11
/product="Glu-tRNA Gln amidotransferase (C subunit)"
/protein_id="BA98212.1"
/db_xref="GI:8978375"
/translation="MEQPHLDREILLAKASALQSEELIQEYQTSLSAVITSKEA
LAIEIDDDSCSESLFMHVNVNEDLREDSVTSDFNREELRWVESLGLVYPAVAK"
895..2373
/gene="gata"
895..2373
/gene="gata"
895..2373
/gene="gata"
895..2373
/codon_start=1
/transl_table=11
/product="Glu tRNA Gln amidotransferase (A subunit)"
/db_xref="GI:8978376"
/translation="MRTSALEAKAVTLGELTATGYQHPPHRIEAEQYAFISL
CKEQLAEALIDKKRSRGEPLGKLAGVPCIKNDIHVTGLTKTCAVLEVDPEPD
ATVERIKKEGIIIGKLNMBEFAMGSTTLVSAPHPNPDLSRVPGSSGGSAAY
SARCPVALGSDTGGISIRQPAFCGVYFKEYSRKGIVAFASIDQIGPLANTY
EDVALMNVFSGRDKDTSREPFDSMSKLSIPEVPRGIVPRFLEGLDRENF
FSSLAIPGESETHLVDELIDLSHAIVSYIYILASAEATNLARDGVYGYCRSQAH
ISQLYDSRGSGFGKEVNRKILLGNIVYLSARQNVYIKKAAVRAKIKARITPERK
EILAMPVCSSPAFTGELIDPVTLIDLYIVANMLAYLAIVAPSGFSKGLPLGLQ

```



gene  
CDS

GP:445586; identified by sequence similarity:  
putative:polymorphic membrane protein G family, authentic  
frameshift"  
complement(9393. .11918)  
/gene="CP0761". .11918)  
complement(9393. .11918)  
/gene="CP0761". .11918)  
/note="similar to GP:4376270; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="polymorphic membrane protein G family"  
/protein\_id="AF38561.1"  
/db\_xref="GI:7189674"  
/translation="MKIPLEFLISLVPLTSMNLGAAATEELASNSPDGTTSTTS  
FSKTSATDGTCTVYKDSVINEVPKGTOSTGKFNAAAGDINFGGSGSPFS  
NIDATTSACATSEPAKNTVTISGFSALSFKSPASTYNGGAIINVKNNLSGLTD  
KVLIDNEFTGDDGAIACGSLKIANKSLISFGNSSSTRGGA1HKKNLISLGGTEL  
PQNTAPTAAGKGAIAIADSGTSLISGSDGDIIPBNTIGATGVYSHSA1IDGTSK  
ITALRAOCHTIFYDPIYVGTSTVADALINSPDGNK EYTGIVFSGEKLTEAE  
AKDEKNTSKLQNVAFKNGTVYKGVYLSANGFSDAKSLIMDGLSVNTESI  
ELNLEINLIDSKKRIKLSAATAKDIIDIPVYLISDESPONGPNEHPSYDQ  
IIEIDAKDQIVISADRSIDAVOSPFGYQKWTINNSTDKKATVSAKOSENPTEAQ  
EAPLVNLMGSEFIDVRSFQNFTELEGAPYERKRWAGISVLRHSGENQKTRH  
VSGAVVGASTRMFGDITSLGFQDLFARBDKFMNTAKYAGSLRLOHDSLYSV  
VSLIEBGGIRELILPYVSKLPCSEYGGSLYGHTRMKTESLPPPLSLSDHSM  
GGYVWAGELGTREVAVENTSGRPFQETPVKQAYARSDSEVELGALS1RDSDSHL  
YNATPLGKIKERPAQYHYVAAASPDVCRNPCTTLLISNOSMKTGKSNLARO  
AGIVQASGRFSAAELEFNGNPFEMRSRSTNVMDGSKTKF"

BASE COUNT 3532 a 2722 c 2431 g 3441 t 1 others  
ORIGIN

Query Match 99.3%; Score 2508.4; DB 1; Length 12127;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2515; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 ATGAAGATTCACATCCGCTTTTATGATATCATATTACCTACGCTTTCATGCAAT 60  
DB 11918 ATGAAGATTCACATCCGCTTTTATGATATCATATTACCTACGCTTTCATGCAAT 11859  
QY 61 TTATTAGAGCTGCTACTACCGAAGAGCTATCGGCTAGCAATAGCTTTCATGCAATCA 120  
DB 11858 TTATTAGAGCTGCTACTACCGAAGAGCTATCGGCTAGCAATAGCTTTCATGCAATCA 11799  
QY 121 TCACACACAGCTTTTCTAGTAAACATCATCGGCTACACATGGCCACCATTAAGTTT 180  
DB 11798 TCACACACAGCTTTTCTAGTAAACATCATCGGCTACACATGGCCACCATTAAGTTT 11739  
QY 181 AAGATTCTGATTTATAGAAAATGTACCCAAAACAGGGAACTCAGTCTACTAGTTGT 240  
DB 11738 AAGATTCTGATTTATAGAAAATGTACCCAAAACAGGGAACTCAGTCTACTAGTTGT 11679  
QY 241 TTTAAAAATGACGCTGAGCTGAGATCTAAATTTCTTAAAGAGGGGATTTTCTTTCACA 300  
DB 11678 TTTAAAAATGACGCTGAGCTGAGATCTAAATTTCTTAAAGAGGGGATTTTCTTTCACA 11619  
QY 301 TTTAGCAATTCATGATGCAACACGCTTTCGAGCTCTATTGGAATGGAAGCATTAAT 360  
DB 11618 TTTAGCAATTCATGATGCAACACGCTTTCGAGCTCTATTGGAATGGAAGCATTAAT 11559  
QY 361 AAGACAGTCACTTATCAGATTTTTCGCACTTTCTTTTAAATCCCAAGCAAGTACA 420  
DB 11558 AAGACAGTCACTTATCAGATTTTTCGCACTTTCTTTTAAATCCCAAGCAAGTACA 11499  
QY 421 GTGACTAATGATTTGGAGCTATCATATGTTAAAGGAATTTAAAGCTATTGATGAT 480  
DB 11498 GTGACTAATGATTTGGAGCTATCATATGTTAAAGGAATTTAAAGCTATTGATGAT 11439  
QY 481 AAGGTATTTGATGAGCAATTTCTCAACAGGAGATGCGGACCAATTAATTGTGACGGC 540  
DB 11438 AAGGTATTTGATGAGCAATTTCTCAACAGGAGATGCGGACCAATTAATTGTGACGGC 11379

QY 541 TCCTTGAAGATCGCAACAAATTAAGTCCCTTTCTTTATGGAATTAAGTCTTCACACGT 600  
DB 11378 TCCTTGAAGATCGCAACAAATTAAGTCCCTTTCTTTATGGAATTAAGTCTTCACACGT 11319  
QY 601 GCGGAGCAGATTTCTACCAAAAACCTCACACATCTCTCTGTGGGAAACTCTATTTCAG 660  
DB 11318 GCGGAGCAGATTTCTACCAAAAACCTCACACATCTCTCTGTGGGAAACTCTATTTCAG 11259  
QY 661 GGAATACAGCGCTACGGCTGCTGTTAAAGAGAGTCTATCCGATTTGACAGACTGCG 720  
DB 11258 GGAATACAGCGCTACGGCTGCTGTTAAAGAGAGTCTATCCGATTTGACAGACTGCG 11199  
QY 721 ACCCTATCCATTTCTGAGACAGTGGGACATTAATCTTTAAAGCAATTAAGAGAGCT 780  
DB 11198 ACCCTATCCATTTCTGAGACAGTGGGACATTAATCTTTAAAGCAATTAAGAGAGCT 11139  
QY 781 ACAGGAACCGTCTCTCATAGTGTATTGATTAGGAACCTAGCGTAAAGATTAAGTCT 840  
DB 11138 ACAGGAACCGTCTCTCATAGTGTATTGATTAGGAACCTAGCGTAAAGATTAAGTCT 11079  
QY 841 CGTCTCGGAGAGACATAGATTAATCTTTTATGATCCGATTTACTGTAACAGATCGACA 900  
DB 11078 CGTCTCGGAGAGACATAGATTAATCTTTTATGATCCGATTTACTGTAACAGATCGACA 11019  
QY 901 TCTGTCTGATGCTCTCAATATTATAGCCCTGATCTGAGATTAACAAAGATATACG 960  
DB 11018 TCTGTCTGATGCTCTCAATATTATAGCCCTGATCTGAGATTAACAAAGATATACG 10959  
QY 961 GGAACCATTAAGTCTTTCTGAGAGACGCTACGAGGACAGAGCTAAAGATGAGAGAC 1020  
DB 10958 GGAACCATTAAGTCTTTCTGAGAGACGCTACGAGGACAGAGCTAAAGATGAGAGAC 10899  
QY 1021 CGCAGCTCTAATTAATTAATCTTCAAAATGTTGCTTTTAAATGGAGCTAGTTTAAAGGT 1080  
DB 10898 CGCAGCTCTAATTAATTAATCTTCAAAATGTTGCTTTTAAATGGAGCTAGTTTAAAGGT 10839  
QY 1081 GATGCTGTTTAAATGAGCAAGGTTTCTCTCAGAGTCAAACTCTAAGTTGATTATGAT 1140  
DB 10838 GATGCTGTTTAAATGAGCAAGGTTTCTCTCAGAGTCAAACTCTAAGTTGATTATGAT 10779  
QY 1141 TTAGGAGCGTCTGTTGTTGCAAAACCCGAAAGTATCGAGTTAAAGCAATTTGGAATTAAT 1200  
DB 10778 TTAGGAGCGTCTGTTGTTGCAAAACCCGAAAGTATCGAGTTAAAGCAATTTGGAATTAAT 10719  
QY 1201 ATGAGCTCTCAAGGAAGGGAAGAAAGATTAAGAACTCGTCTGCAAGCTCAAGAAAT 1260  
DB 10718 ATGAGCTCTCAAGGAAGGGAAGAAAGATTAAGAACTCGTCTGCAAGCTCAAGAAAT 10659  
QY 1261 ATTGATATGATGCTCTGTTGTACTGGCAATTAGCGATGAGATTTTATCAAAATGGC 1320  
DB 10658 ATTGATATGATGCTCTGTTGTACTGGCAATTAGCGATGAGATTTTATCAAAATGGC 10599  
QY 1321 TTTTGAATGAGACATTTCTTATGATGGATTTCTTGAATGATGCTGGGAAAGCATC 1380  
DB 10598 TTTTGAATGAGACATTTCTTATGATGGATTTCTTGAATGATGCTGGGAAAGCATC 10539  
QY 1381 GTGATTTCTGACAGATTTCTGCAATTAATTAAGTCAATCTCGATVAGGCTTTCAGGGA 1440  
DB 10538 GTGATTTCTGACAGATTTCTGCAATTAATTAAGTCAATCTCGATVAGGCTTTCAGGGA 10479  
QY 1441 AAGTGAACATTAATGATGATGATGATTAAGAAAGCTACGCTTCTTGGGAAAGCA 1500  
DB 10478 AAGTGAACATTAATGATGATGATGATTAAGAAAGCTACGCTTCTTGGGAAAGCA 10419  
QY 1501 AGTTTAAATCCACGCTGAGAGAGAGGCTCGTATGTTCTTATCTTCTTGGGATTC 1560  
DB 10418 AGTTTAAATCCACGCTGAGAGAGAGGCTCGTATGTTCTTATCTTCTTGGGATTC 10359  
QY 1561 TTTATGATGTTGCTGCTCCCAAAATTTTATAGAGTAGTACTGAAAGTGTCTCTTAC 1620  
DB 10358 TTTATGATGTTGCTGCTCTTCCAGAAATTTTATAGAGTAGTACTGAAAGTGTCTCTTAC 10299  
QY 1621 GAAAGAGATTTTGGGTTGAGGACATTTCCAAATGTTTGCATAGAGCGGCTCTGAAAT 1680

QY 1441 AAGTGACATCAATGTCCTACTGATGATGAAAGACGATTCCTGGCAAGCAA 1500  
 DB 9318 AAGTGACATCAATGTCCTACTGATGATGAAAGACGATTCCTGGCAAGCAA 9377  
 QY 1501 AGTTTAATCCACGCTGAGAGAGGCTCGTGTCTAATCTCTGGGCTTCT 1560  
 DB 9378 AGTTTAATCCACGCTGAGAGAGGCTCGTGTCTAATCTCTGGGCTTCT 9437  
 QY 1561 TTTATAGATGTCCTGCTCCAAATTTATAGAGTACTGAAGTCTCTCTAC 1620  
 DB 9438 TTTATAGATGTCCTGCTCCAAATTTATAGAGTACTGAAGTCTCTCTAC 9497  
 QY 1621 GAAAGAGATTTTGGGTGAGGATTTCCAAATTTTTCATAGAGCGGTCTGTA 1680  
 DB 9498 GAAAGAGATTTTGGGTGAGGATTTCCAAATTTTTCATAGAGCGGTCTGTA 9557  
 QY 1681 CAAGAGATTTTCCGCTGATGATGAGGATGCTGTAGTGTCTAGCAGAGATCCG 1740  
 DB 9558 CAAGAGATTTTCCGCTGATGATGAGGATGCTGTAGTGTCTAGCAGAGATCCG 9617  
 QY 1741 GGTGTGATACCTGCTCTGCTGGTGTGCTGAGCTTTCGCGTGAAGACTCTT 1800  
 DB 9618 GGTGTGATACCTGCTCTGCTGGTGTGCTGAGCTTTCGCGTGAAGACTCTT 9677  
 QY 1801 ATGAATACCAATTTCCGCAAGACCTACGAGATCTTACGTTGACAGACATGCTTC 1860  
 DB 9678 ATGAATACCAATTTCCGCAAGACCTACGAGATCTTACGTTGACAGACATGCTTC 9737  
 QY 1861 CATATCTGCTGCTGATCTCTTTAGAGAGGAGGACTCCGAGATCTCTGCTT 1920  
 DB 9738 CATATCTGCTGCTGATCTCTTTAGAGAGGAGGACTCCGAGATCTCTGCTT 9797  
 QY 1921 TATGTTTCCAGACTCTGCTGCTCTTCTATGAGGACACTTACGACGCTACATG 1980  
 DB 9798 TATGTTTCCAGACTCTGCTGCTCTTCTATGAGGACACTTACGACGCTACATG 9857  
 QY 1981 CATGCTATGAGACGAGTCTCTACCCCCCCCCCGAGCTCTGACGATCATCT 2040  
 DB 9858 CATGCTATGAGACGAGTCTCTACCCCCCCCCCGAGCTCTGACGATCATCT 9917  
 QY 2041 TCTGTGGGAGGATGCTGCTGGCTGAGAGCTGGAGACTGCTGTGAAATAC 2100  
 DB 9918 TCTGTGGGAGGATGCTGCTGGCTGAGAGCTGGAGACTGCTGTGAAATAC 9977  
 QY 2101 AAGCGAGAGATTTTCCGAGAGTACATCCATTTGTTAAAGTCCAAAGCTTTACTCG 2160  
 DB 9978 AAGCGAGAGATTTTCCGAGAGTACATCCATTTGTTAAAGTCCAAAGCTTTACTCG 10037  
 QY 2161 GCGCAAGATAGCTTGTGTAAGTACAGAGTATCAGTGTGATTTAGTATCGACTCT 2220  
 DB 10038 GCGCAAGATAGCTTGTGTAAGTACAGAGTATCAGTGTGATTTAGTATCGACTCT 10097  
 QY 2221 TATTAACCTGCAATCTCTGTAAGTACAGTATGAAAGCTTTGCAAGATTTAT 2280  
 DB 10098 TATTAACCTGCAATCTCTGTAAGTACAGTATGAAAGCTTTGCAAGATTTAT 10157  
 QY 2281 CATGTTGATGATGATCTCCAGATGTTTCTGATGATGATGATGATGATGATGAT 2340  
 DB 10158 CATGTTGATGATGATCTCCAGATGTTTCTGATGATGATGATGATGATGATGAT 10217  
 QY 2341 CATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400  
 DB 10218 CATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10277  
 QY 2401 ATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460  
 DB 10278 ATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10337  
 QY 2461 GCGCTTGAATGCGGGGATCTCTGATGATGATGATGATGATGATGATGATGAT 2520  
 DB 10338 GCGCTTGAATGCGGGGATCTCTGATGATGATGATGATGATGATGATGATGAT 10397  
 QY 2521 TTTTAG 2526

DB 10398 TTTTAG 10403  
 RESULT 4  
 AE002235/c  
 LOCUS  
 DEFINITION  
 Chlamydia pneumoniae AR39, section 63 of 94 of the complete genome.  
 ACCESSION  
 AE002235  
 VERSION  
 AE002235.2  
 KEYWORDS  
 GI:8163495  
 ORGANISM  
 Chlamydia pneumoniae AR39.  
 SOURCE  
 Chlamydia pneumoniae AR39.  
 REFERENCE  
 1 (bases 1 to 12127)  
 Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Kouri, H., Craven, B., Bowman, C., Dodson, R., Gwin, M., Nelson, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.  
 Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39  
 Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
 2 (bases 1 to 12127)  
 Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Kouri, H., Craven, B., Bowman, C., Dodson, R., Gwin, M., Nelson, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.  
 Direct Submission  
 Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
 On Jun 1, 2000 this sequence version replaced gi:1189672.  
 LOCATION/Qualifiers  
 1..12127  
 /organism="Chlamydia pneumoniae AR39"  
 /strain="AR39"  
 /db\_xref="taxon:115711"  
 /note="synonym: Chlamydia pneumoniae AR39"  
 /complement(150..2987)  
 /gene="CP0757"  
 /note="this region contains an authentic point mutation, causing a premature stop, and is not the result of a sequencing artifact; similar to GP:445586; Identified by sequence similarity; putative:polymorphic membrane protein G family authentic point mutation"  
 /complement(3003..3191)  
 /gene="CP0758"  
 /complement(3003..3191)  
 /note="CP0758"  
 /gene="hypothetical protein; identified by Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=1  
 /product="hypothetical protein"  
 /protein\_id="AF38560.1"  
 /db\_xref="GI:7189673"  
 /translation="WRLLSCPMLFIACASFQFQEMOGRNIOISDANNSLGELE  
 SISTRGVSCLEIAREIAR"  
 /complement(3336..6086)  
 /gene="CP0759"  
 /note="this region contains a gene with one or more premature stops or frameshifts, and is not the result of a sequencing artifact; similar to GP:445586; Identified by sequence similarity; putative:polymorphic membrane protein G family, degenerate"  
 /complement(6325..9110)  
 /gene="CP0760"  
 /note="this region contains an authentic frame shift and is not the result of a sequencing artifact; similar to

```

CDS
5938..7509
/gene="Cp0012"
/codon_start=1
/transl_table=11
/protein_id="A01817.1"
/product="hypothetical protein"
/translation="MYOENLRLELRLYNSVOKSYADRLFSYEKTKVHDTPILPME
DREKCAEAEKFLDOKILLDYGKSTFPMLENEININDPMSGILNVRFRKYFOEVD
DSEKRNKYLIOKLEDDYKLLDESSKESLEANKLLSDVBLEADKTFPLAKODE
VETVRKDLARKIGTVDPKODTEAKKVELESLDFTDISELVOCLDQDITWKE
QDVKDLARIDOLEODLEAKREAEDELRLSLNLRKSKTMDLRANKHINADSTW
MISOIEKMDKARKIKILEDITVLEIDELTCLSELPLTRILRLYSYLKFTI
CSELLKMTSEVENNIYVOEYVLOLAGEKLOGISQRFEGKODDFEALNEEOVALOK
RLBELONFEIOGFNFKEDFKAADLYIRSTAEOKMNDVCOMELFRYHYEVNPD
LLELMNCADSYRDARKKCLSLDEKELLOKKEEFYQKQQRHADRNSRHTYOK
LRLAELELELKKI"
7878..10403
/gene="pmp_2"
/note="Cp0013"
7878..10403
/gene="pmp_2"
/codon_start=1
/transl_table=11
/product="Polymorphic Outer Membrane Protein G Family"
/db_xref="GI:4376270"
/translation="MKIPLRFLISLVPLISMNLGAATTEELSAISNFDGTTSTTS
FSKRTSATDGINYVERKSVLENPVGTSTQSCPKNDAAAGDLNFGGSEFSES
NIDATTAAGTAIGSEANKTVLISGFALSFLKSPASTVAGDALNVKNLILND
KVLIONESTDGAINGCAISGLKIANKSLFIGNSGTSGAIIHKNLILSGGEL
FOGNTPTAGKGGALAIADSGTILSGDGLIIFEGNTGATGVSHSIDGTSK
ITLRAAGHTIYEPDPIYVSTSVADALNIPGMKDEYTGITVESELTARE
AKRKNTSKLONVAFKNTVYLKGVLSANSPFODANSKILMDIGTSLVANTESI
ELTNLEINISLNRKIKLSAATADKDIIDRPVLAIDSEFYQGFNEHSDYD
ILBDAKGDIVIASRSDAVOSPYGQKWTINSTDKKAIVSAKSEFPTEDQ
EALVPLNLWGSFIDVRSFQNFTELEGAEYERKFWAGISVNLHSGRENKRFH
VSGAVAYGASTRMRGGDTLSLGFAPLARKDYEMNTNEKTAGSLRLQHDASTAK
VSLLEGGLRELLPYVSKTLPSCSYGOLSYGHTRMKTESLPPPLTSDHISM
GGYVAGELCTRAVNTSRSRFFOETTPRYKQAYARDSVEYLGASIRDSDSL
YINAIPLGILKLRFAEDYIHVVAATSPDCRSNPCTTLLTNSQSKTKGSLNAQ
AGIVQASGEFSLAAALEFNGFEWGRSSRSYNVADASKIKF"
BASE COUNT 3217 a 1747 c 2651 g 2959 t
ORIGIN
Query Match 99.3% Score 2508.4; DB 1; Length 10574;
Best Local Similarity 99.6% Pred. No. 0;
Matches 2515; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 1 ATGAAGATTCCACGCGCTTTTATGATATCATAGTACCTAGCGTTTCATGCGAAT 60
DB 7878 ATGAAGATTCCACGCGCTTTTATGATATCATAGTACCTAGCGTTTCATGCGAAT 7937
QY 61 TTATTAGAGCTGCTACTACGSAAGAGCTATCGGCTAGCAATAGCTTCGATGAACTACA 120
DB 7938 TTATTAGAGCTGCTACTACGSAAGAGCTATCGGCTAGCAATAGCTTCGATGAACTACA 7937
QY 121 TCACAACAAGCTTTTCTAGTAAAAACATCATCGGCTACAGATGGCACCAATATATGTTTT 180
DB 7998 TCACAACAAGCTTTTCTAGTAAAAACATCATCGGCTACAGATGGCACCAATATATGTTTT 8057
QY 181 AAGATTTCTGAGTATAGAAAAATGTACCCAAAAAGGGAATCTCAGTCTACATGTTGT 240
DB 8058 AAGATTTCTGAGTATAGAAAAATGTACCCAAAAAGGGAATCTCAGTCTACATGTTGT 8117
QY 241 TTTAAAAATAGCGCTGAGCTGAGATCTAAATTTCTTAGAGAGGGGATTTTCTTTACA 300
DB 8118 TTTAAAAATAGCGCTGAGCTGAGATCTAAATTTCTTAGAGAGGGGATTTTCTTTACA 8177
QY 301 TTTAGCAATATCATGATGACACACGCGCTTCTGAGCTGCTATTTGGAAGTGAAGCAGCTAAT 360
DB 8178 TTTAGCAATATCATGATGACACACGCGCTTCTGAGCTGCTATTTGGAAGTGAAGCAGCTAAT 8237

```

FEATURES	source
gene	/organism="Chlamydia pneumoniae CML029"
CDS	/strain="CML029"
	/db_xref="taxon:115713"
	/gene="Cpn0008"
	/gene="Cpn0008"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"

[illegible]

gene 14954..17479  
/gene="omp7"  
CDS 14954..17479  
/gene="omp7"  
/note="putative"  
/codon\_start=1

Query Match 99.48; Score 2510; DB 2; Length 26920;  
Best Local Similarity 99.68; Pred. No. 0;  
Matches 2516; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ATGAAGATTCACATCCGCTTTTATGATATCATATGATACCTAGCTTTCTATGTCGAT 60  
Db 14954 ATGAAGATTCACATCCGCTTTTATGATATCATATGATACCTAGCTTTCTATGTCGAT 15013  
QY 61 TTATTAGAGCTGCTACTACCGAAGAGCTATCGCTACGAATAGCTTGATGAGACTACA 120  
Db 15014 TTATTAGAGCTGCTACTACCGAAGAGCTATCGCTACGAATAGCTTGATGAGACTACA 15073  
QY 121 TCACCAACAACCTTTCTAGTAAACATCATCGGCTACAGATGCGACCAATTATGTTT 180  
Db 15074 TCACCAACAACCTTTCTAGTAAACATCATCGGCTACAGATGCGACCAATTATGTTT 15133  
QY 181 AAGATTCCTAGTATAGAAATGTACCCAAACAGGGGAAACTCACTACTAGTTGT 240  
Db 15134 AAGATTCCTAGTATAGAAATGTACCCAAACAGGGGAAACTCACTACTAGTTGT 15193  
QY 241 TTTAAAAATGCGTGCAGCTGAGATCTAAATTTCTTAGAGAGGGATTTCTTTTACA 300  
Db 15194 TTTAAAAATGCGTGCAGCTGAGATCTAAATTTCTTAGAGAGGGATTTCTTTTACA 15253  
QY 301 TTTAGCAATATCATGAGCAACCGCTTCTGAGCTGCTATTTGGAAGTGAAGCAAGCTAAT 360  
Db 15254 TTTAGCAATATCATGAGCAACCGCTTCTGAGCTGCTATTTGGAAGTGAAGCAAGCTAAT 15313  
QY 361 AAGACGTCACGTTATGAGATTTTTCGGCACTTTCTTTCTTAAATCCCAAGCAATACA 420  
Db 15314 AAGACGTCACGTTATGAGATTTTTCGGCACTTTCTTTCTTAAATCCCAAGCAATACA 15373  
QY 421 GTGACTATGATTTGGAGAGCTATCATATGTTTAAAGGAATTTAAGCTTATGATATGAT 480  
Db 15374 GTGACTATGATTTGGAGAGCTATCATATGTTTAAAGGAATTTAAGCTTATGATATGAT 15433  
QY 481 AAGTATTTGATTCAGACAATTTCTCAACAGAGATGGCGGAGCAATTAATTTGCGAGGC 540  
Db 15434 AAGTATTTGATTCAGACAATTTCTCAACAGAGATGGCGGAGCAATTAATTTGCGAGGC 15493  
QY 541 TCCTTGAAGATCGCAAAACAATTAAGTCCCTTTCTTTATTTGGAATAGTCTTCAACAGT 600  
Db 15494 TCCTTGAAGATCGCAAAACAATTAAGTCCCTTTCTTTATTTGGAATAGTCTTCAACAGT 15553  
QY 601 GCGGAGCGATTCATACCAAAAACCTCAACACTATCTTCTGGTGGGAAACTCTATTTTCAAG 660  
Db 15554 GCGGAGCGATTCATACCAAAAACCTCAACACTATCTTCTGGTGGGAAACTCTATTTTCAAG 15613  
QY 661 GGAATATCAGGCGCTACGCGCTGTAAAGAGAGTCTATCGGATTCAGACTCTGGC 720  
Db 15614 GGAATATCAGGCGCTACGCGCTGTAAAGAGAGTCTATCGGATTCAGACTCTGGC 15673  
QY 721 ACCCTATTCATTTCTGAGACAGTGGGACATTAATCTTTGAAGCAATACGATAGAGCT 780  
Db 15674 ACCCTATTCATTTCTGAGACAGTGGGACATTAATCTTTGAAGCAATACGATAGAGCT 15733  
QY 781 ACAGGAACCGTCTCTATAGTCTATTGATTTAGGAACATAGCGCTTAAGATTAAGTGCCTTA 840  
Db 15734 ACAGGAACCGTCTCTATAGTCTATTGATTTAGGAACATAGCGCTTAAGATTAAGTGCCTTA 15793  
QY 841 CGTCTGCGAAGGACATAGATATCTTTTATGATCCGATTAACCGATTAACCGCTTA 900  
Db 15794 CGTCTGCGAAGGACATAGATATCTTTTATGATCCGATTAACCGATTAACCGCTTA 15853  
QY 901 TCTGTTCTGATGCTCTCAATATTAATAGCCCTGATAGTGAATTAACAAAGATATACG 960  
Db 15854 TCTGTTCTGATGCTCTCAATATTAATAGCCCTGATAGTGAATTAACAAAGATATACG 15913

Db 15854 TCTGTTCTGATGCTCTCAATATTAATAGCCCTGATAGTGAATTAACAAAGATATACG 15913  
QY 961 GGAACCATAGCTTTTCTGAGAGAACTCAGAGCAGAAAGCTAAAGATGAAAGAAC 1020  
Db 15914 GGAACCATAGCTTTTCTGAGAGAACTCAGAGCAGAAAGCTAAAGATGAAAGAAC 15973  
QY 1021 CGCACTTCTAATTAATCTCAAAATGTTGCTTTTAAATGAGACTGTATTTTAAAGT 1080  
Db 15974 CGCACTTCTAATTAATCTCAAAATGTTGCTTTTAAATGAGACTGTATTTTAAAGT 16033  
QY 1081 GATGCTGTTTAAAGTGGCAAGCGTTTCTCTAGAGTGCACAACTGAATGATTATGAT 1140  
Db 16034 GATGCTGTTTAAAGTGGCAAGCGTTTCTCTAGAGTGCACAACTGAATGATTATGAT 16093  
QY 1141 TTAGGAGCTGTTGTTGCAACACGGAAGTATCGAGTTAAGAAATTTGGAATTAAT 1200  
Db 16094 TTAGGAGCTGTTGTTGCAACACGGAAGTATCGAGTTAAGAAATTTGGAATTAAT 16153  
QY 1201 ATGACTCTCAGCAACGCGGAAAGATTAATACTAGTCTGCGACAGCTCAGAAAGAT 1260  
Db 16154 ATGACTCTCAGCAACGCGGAAAGATTAATACTAGTCTGCGACAGCTCAGAAAGAT 16213  
QY 1261 ATTCGTATAGATCGCTGTTGTATCTGCAATTAGCGATGAGAGTATTATCAAAATGCG 1320  
Db 16214 ATTCGTATAGATCGCTGTTGTATCTGCAATTAGCGATGAGAGTATTATCAAAATGCG 16273  
QY 1321 TTTTGAATGAGCAACCTCTATGATGAGATTTCTTGAATGATGCTGGGAAAGATC 1380  
Db 16274 TTTTGAATGAGCAACCTCTATGATGAGATTTCTTGAATGATGCTGGGAAAGATC 16333  
QY 1381 GTGATTTCTGAGATTCCTGCGAGTATTAATCTGTACAAATCCGATATGCTATCAGGA 1440  
Db 16334 GTGATTTCTGAGATTCCTGCGAGTATTAATCTGTACAAATCCGATATGCTATCAGGA 16393  
QY 1441 AAGTGACATCAATTTGCTACTGATGATTAAGAAAGCTACGAGTTCTTGCGCAACCAA 1500  
Db 16394 AAGTGACATCAATTTGCTACTGATGATTAAGAAAGCTACGAGTTCTTGCGCAACCAA 16453  
QY 1501 AGTTTAAATCCACTGCTGAGCAGAGAGGCTCCGTTACTTCTATCTTTTGGGGTCT 1560  
Db 16454 AGTTTAAATCCACTGCTGAGCAGAGAGGCTCCGTTACTTCTATCTTTTGGGGTCT 16513  
QY 1561 TTTATAGATGCTCTCCCTTCCAAATTTTATAGCTAGGATGATGAGGTGCTCTTAC 1620  
Db 16514 TTTATAGATGCTCTCTCTTCCAAATTTTATAGCTAGGATGATGAGGTGCTCTTAC 16573  
QY 1621 GAAAGAGATTTGGTTCAGAGGCTATTCAGATTTTGCATAGAGCGGCTGTAAAT 1680  
Db 16574 GAAAGAGATTTGGTTCAGAGGCTATTCAGATTTTGCATAGAGCGGCTGTAAAT 16633  
QY 1681 CAAGGAATTCCTCATGAGAGAGGAGGCTGTAGAGTGTCTACAGAGAGATGCGC 1740  
Db 16634 CAAGGAATTCCTCATGAGAGAGGAGGCTGTAGAGTGTCTACAGAGAGATGCGC 16693  
QY 1741 GGTGATATACCTTGTCTGAGGTTTGTCTAGCTCTTTGCGGCTGCAAAAGACTACTTT 1800  
Db 16694 GGTGATATACCTTGTCTGAGGTTTGTCTAGCTCTTTGCGGCTGCAAAAGACTACTTT 16753  
QY 1801 ATGAATACCAATTTGCAAAAGACTACAGAGATCTTTAGCTTTGCGAGCATGCTTCC 1860  
Db 16754 ATGAATACCAATTTGCAAAAGACTACAGAGATCTTTAGCTTTGCGAGCATGCTTCC 16813  
QY 1861 CTATACCTGTTGGTGAATATCTTTTGAAGAGGAGAGACTCCCGGAGATCTGTTGCT 1920  
Db 16814 CTATACCTGTTGGTGAATATCTTTTGAAGAGGAGAGACTCCCGGAGATCTGTTGCT 16873  
QY 1921 TATGTTTCAAGACTGCGCGTCTTTCTATGAGGAGCTTACCTACGCGCATACGAT 1980  
Db 16874 TATGTTTCAAGACTGCGCGTCTTTCTATGAGGAGCTTACCTACGCGCATACGAT 16933  
QY 1981 CATGCAATGAAGACGAGTCTTACCCCGCCCGCCGAGCGCTCTGAGCGATCATACT 2040  
Db 16934 CATGCAATGAAGACGAGTCTTACCCCGCCCGCCGAGCGCTCTGAGCGATCATACT 16993

PTQSPGSHLVLDLGTKLIAKEDIAITGLAIDIDLSLSSSTAAVIAKANTANKOISVTD  
 STELISPTGNATEDLRMRNSQFFPLLSLEPGGSGVTVTAGLFPVSGQGMKL  
 AMTGNGKVGEEFMKINRKPRPEKNGINVLPIILMGNAVNRSLMOVETHASQLOTD  
 RGTWIDGICGNPFHVASSEDNIRYRNHSGYIVLSVNNETPKRYTSMASQLESRKDY  
 AVSNBETRYLASYLTQYTTSLGNTFRASRNPNVAGILSRFLONPLMTFHLICAY  
 GHATNDMTDYANFPMVKNMWRNMCWALIECGSGMPLLVFENGRLFOGAIPKMLDOLVY  
 AVQGFKEFTTADGRRFSNGSLTISVPLGIREKALISODVLYDFSFSYIPIDIFKPD  
 SCEALVLSIGDSMWLPAAVSHRAVFGSGTGRHNDYTELLCRGSIICRPHRMYNI  
 NCGSKFRF"  
 3976. 6870  
 /gene="hb1"  
 /db\_xref="db1"  
 3976. 6870  
 /gene="hb1"  
 /note="hypothetical"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HB1 protein"  
 /protein\_id="CAB37076.1"  
 /db\_xref="GI:445892"  
 /db\_xref="SPTREMBL:Q923C8"  
 /translation="MRKLRLAIVLIALISIIILAGGVLLTVAIPGLSSVSPAGMG  
 ACAGCVMLAIGIDVLLKRRVPIYLASVTTPTGSPRSGISISGADSTIRSLPYTL  
 IDEGHPQMKRLITAIIVFSITILASGVLLTVAIPGLSSVSPAGMACLAGC  
 VMLAGIDVLLKRRVPIYLASVTTPTGSPRSGISISGADSTIRSLPYPLDGH  
 QSMRLRLAIVLIVFSITILASGVLLTVAIPGLSSVSPAGMACAGCVMLAIG  
 IDVLLKRRVPIYVAPVPEEVYIDIDIESIRLOQEAALALRPEEMSAFEGYIKV  
 VESHLNNKSLPYDGHLEKTKHOIRVRSSTKAMVPELDIRIRIPPEEPPFISAR  
 KRILDTATVPERKITLTOLEENNRKAFSYIKOSITIKITIDNEKILAKRMILSKS  
 ICRPTITENHEGVAKSLHKNVLRVFLALSDENVDPSPDKKMDLGIIPORDLPE  
 SLEDRKKTIMKEHAMLESLSRYRFLALSDENVDPSPDKKMDLGIIPORDLPE  
 ISRDQOMKALHKEQSLYQARDRLDOSKRNQKELKAEQEIISMEVSKFEI  
 ERVORAIQIKLYPNILEREETTGQETVPTVGTASDLTDILGRIEVSKREDI  
 ONOESCVAIVLSHEVMSMEVQKQEPKKEFDODMGSLERPFTHIELEVLQDYS  
 KHLSEKRVNNKKEVOYAKFRLKYLESDLEGITLAQTESASILLTOELPIIATRALE  
 KAVFGSICCALASAKPYFEEDPRFQSDTQRLTLTLRLQAKASLEBEIKRSENLE  
 NDIAERRLKESQTFEPAGLGLREIAVESTYDLRSLTNTWEGTPESEKPYFSMYL  
 NYINEKRRAKRLVEMTORVRFKMALEAMQFNEALQLELSIQAPSE"  
 7150. 8058  
 /gene="hb2"  
 7150. 8058  
 /gene="hb2"  
 /note="hypothetical"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HB2 protein"  
 /protein\_id="CAB37077.1"  
 /db\_xref="GI:445893"  
 /db\_xref="SPTREMBL:Q925Q3"  
 /translation="MQVFLSPQLPPPRRLSGSCSKRLSTILLVLVGLLLTGLIP  
 GLTAGISFGAGISGALGVLYISGLFLVAREVPTVASEEIPRGVSVTPSEEBALE  
 KAKQEPETKILDRLPKELDOLDYIOEVFACLERLKDPRKYEDRGLTEAKKLEAFD  
 VYKDMSEPLDIOVRNBEAVYVHCCDPLENINYEIJSSELRDYICAGCGILPS  
 GDARDLRKSYKEVMDRPMRYTKSMASVWLDSTYGARLFRKANGVLEESYKI  
 LFKSTRDAFYEEERKAKIQDGRKMWL"  
 8185. 9492  
 /gene="hb3"  
 8185. 9492  
 /gene="hb3"  
 /note="hypothetical"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HB3 protein"  
 /protein\_id="CAB37078.1"  
 /db\_xref="GI:445894"  
 /db\_xref="SPTREMBL:Q923C0"  
 /translation="WMLDRYADKPILRKEKEMERHELPHATVYRKASGHAYAKAKA  
 FEKESNENQKRVKVERKWSLGLAEFRNQESRRARERLQTLTYPEVSEERVLER  
 OTRKVNLENLYADIEKRYHRCVROQEHYKWEKKEAYRENGEKVSAEBSVCLQ  
 RLIEDLETWSEKLTIAEESVPEMKDATERKGNKVLSDYTNRLTEICEDAEEMIRIE  
 EEMTLRMVLEPLTMNTPEKASLOYSCKEMLAKVPOCKESPITYRSSQRLRLN  
 ODIOATYVNCORLGDSLDSEKVRCTGDHLREOMKHFVQGLNFINEELMWGAELE  
 TOARLDIATVYPMYEFYLOYHNIRKEKYKVSOMKATBERYRELROAFQGVKMKDLAED  
 TILKEEDYWLADWMLNDRKRNORRLICNKITAAQQRVKG"

9805. 10698  
 /gene="hb4"  
 9805. 10698  
 /gene="hb4"  
 /note="hypothetical"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HB4 protein"  
 /protein\_id="CAB37079.1"  
 /db\_xref="GI:445895"  
 /db\_xref="SPTREMBL:Q923B1"  
 /translation="MLLISGALFTLGLIPGLSAISFGLIGLSALGVLMISGLIC  
 LLVKEIIPVPEEIPREGVSLAPSEPALQAAQKTLAQPKELDOLDTDIDQVFCRL  
 KLDKSESRSFPLNDARKRELRFVDFVEDTLEIFELRLQIVAOEGMDLNLINGRSL  
 KMTASESIDLPHVSKRLGLIPSGVGRGGLKSKKEIYARLMSHCEIHKYAVADR  
 NSYAAEKRAFALCALDESVYRSITQSRDKFLESEAKITMWNHITWLRDANSQC  
 AEKRLGMPNRVNRNLGKQSG"  
 10776. 12119  
 /gene="hb5"  
 10776. 12119  
 /gene="hb5"  
 /note="hypothetical"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HB5 protein"  
 /protein\_id="CAB37080.1"  
 /db\_xref="GI:445896"  
 /db\_xref="SPTREMBL:Q929G3"  
 /translation="MDELTFHELYEKTTFIKRLHKKALAKTTPEKKRSKYLQAVEE  
 ANARLKYVRWDYDQFOAGERLEKLAHYEVSISIRENKTQFTKSLKAYALIE  
 EHYNRCCVROEDYKVEEKREARFERNGKILSPLEBESSLQFPHGLNFSEKLMEL  
 BEHILKQVIAEAEVNEKILSDAESRLLEIVDVKEMPCRILEIKETLMAELPLPT  
 KKAEPKASQYNSKALEMEKYKPKESLAVYSKERVLSLDEDDLRATVBCOKRROG  
 DSGLESEVACAREOLRERIOEFTQGLIVEKEILLCVSRLNTEPCDDGVYKRAP  
 GKFTAQYIDELTYRRVQSRMNTMSERLREGVQACNKLKALSDSDYIKAEETWLY  
 REERNKKEKRLGIVANQRIQIEOPSDIVESSNEKYSIMDKARFLFNRRDHS"  
 12262. 12987  
 /gene="hb6"  
 12262. 12987  
 /gene="hb6"  
 /note="hypothetical"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HB6 protein"  
 /protein\_id="CAB37081.1"  
 /db\_xref="GI:445897"  
 /db\_xref="SPTREMBL:Q923B2"  
 /translation="MTAAPAILHVSPTPEETKVIYPKDSKRALGITLLVIGILLVY  
 CGAIVLSGVISGLSALIVCGLISITSLGVLFVGLILLKRRRELTLEQIBAKQIAE  
 TPADELKLEMYIQSTEKSLERIEGSRVSDQGLNRAQOKILDESSLSITSEFRDL  
 RQLPPEKEITELSGRLLEFIANFQKGRDYYLNLGNLADIRAYWGPNNYKVAMVIE  
 KAKAVVHEITVITMARLEPFP"  
 13014. 14585  
 /gene="hb7"  
 13014. 14585  
 /gene="hb7"  
 /note="hypothetical"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HB7 protein"  
 /protein\_id="CAB37082.1"  
 /db\_xref="GI:445898"  
 /db\_xref="SPTREMBL:Q926P3"  
 /translation="MYOENLRLERLLYNSVOKSYADRLEFSEYKTKMVADTPLPME  
 DKRCAPAKAFLEOQKILLDYGKSLFPMNENDEINLNDPMGMLNTVTRKVPQED  
 DSEKRNHLYLOKLDENYENKLEBSSKSTENANKLISLVYRLDARKKPLFKQOE  
 VETRYKDLARQGTVDPRQDTEARKKVELSLTFPDSITSELVOCLENDQITWKE  
 ODVKLARKLQEGVDDIEAKREAEADLSELRKLSKTYMLDRAKWHLENDQITW  
 WTSQJEMKMKARLKLILKEDITSVLEIDEITCLESLEPLITTRRELTKSYLAKFI  
 CSETLQEMKSVENNIIYOEYQONLGFKIQGISORGKQODDFANLEBOVALQK  
 RLRELTQNEIOGFNMKEDPFKAARDLYIRSTABOKMVDVPCMLFRYHSEYVNP  
 LLELTALNADSVRDYKDKKLSRLDEKELQKEIKKEEYQAKQQRHADRSHHTYYQK  
 LRIAEALTEKKKI"



Db	1141	TTTTAGGACGTGCTGGTGGTCCAAACCGGAAAGTATCCAGTTTAACGAATTTTGGAATTAAT	1200
Qy	1201	ATAGACTCTCTCAGGAACGGGAAAAAGATAAAACTCAGTGCCTGCACAGCTCAGAAAAT	1260
Db	1201	ATAGACTCTCTCAGGAACGGGAAAAAGATAAAACTCAGTGCCTGCACAGCTCAGAAAAT	1260
Qy	1261	ATTGCTATATAGATCCGTCCTGTTGTACTGGCAATTATGCGATATGAGATTTTATCAAAATGGC	1320
Db	1261	ATTGCTATATAGATCCGTCCTGTTGTACTGGCAATTATGCGATATGAGATTTTATCAAAATGGC	1320
Qy	1321	TTTTGAATGAGGACCAATCCATATGATGGAGATTTGTAGATTAGATGGCTGGGAAAGACATC	1380
Db	1321	TTTTGAATGAGGACCAATCCATATGATGGAGATTTGTAGATTAGATGGCTGGGAAAGACATC	1380
Qy	1381	GTGATTTCTGCAGATTTCTGCGAGTATTAATATGCTGTACATCTCCGTATGCGTATCAGGGA	1440
Db	1381	GTGATTTCTGCAGATTTCTGCGAGTATTAATATGCTGTACATCTCCGTATGCGTATCAGGGA	1440
Qy	1441	AAGTGGACAATCAATTGGTCTACTGATGATATAGAAACCTACAGGTTTCTTGGGCAAGCAA	1500
Db	1441	AAGTGGACAATCAATTGGTCTACTGATGATATAGAAACCTACAGGTTTCTTGGGCAAGCAA	1500
Qy	1501	AGTTTATPCCACAGCGCTGAGCAGAGAGGCGCCGTATGTTCCATATCTTTGGGGTTC	1560
Db	1501	AGTTTATPCCACAGCGCTGAGCAGAGAGGCGCCGTATGTTCCATATCTTTGGGGTTC	1560
Qy	1561	TTTATATAGATGTTCCGTCCTCCCAAAATTTTATAGAGCTAGTACTGAAGTGCCTTTAC	1620
Db	1561	TTTATATAGATGTTCCGTCCTCCCAAAATTTTATAGAGCTAGTACTGAAGTGCCTTTAC	1620
Qy	1621	GAAAGACATTTTGGGTGCGACGCAATTTTCCAAATTTTTCATAGAGAGCGGTCTGAAAT	1680
Db	1621	GAAAGACATTTTGGGTGCGACGCAATTTTCCAAATTTTTCATAGAGAGCGGTCTGAAAT	1680
Qy	1681	CAAAAGAAATTCGCGCATGATGATGAGAGAGTCTGTAGTAGTGTCTACAGAGAGATCCG	1740
Db	1681	CAAAAGAAATTCGCGCATGATGATGAGAGAGTCTGTAGTAGTGTCTACAGAGAGATCCG	1740
Qy	1741	GGTGTGATACCTTGCTCTGGGTTTGTCTACGCTCTTTGCGCGTGCAAAAGACTACTTT	1800
Db	1741	GGTGTGATACCTTGCTCTGGGTTTGTCTACGCTCTTTGCGCGTGCAAAAGACTACTTT	1800
Qy	1801	ATGAATATACCAATTTTCGCAAAACCTACGAGAGATCTTACGTTTGCACGACATGCTTCC	1860
Db	1801	ATGAATATACCAATTTTCGCAAAACCTACGAGAGATCTTACGTTTGCACGACATGCTTCC	1860
Qy	1861	CTATPCTGTGGGATGATCCTTTTAGAGAGAGGAGAGACTCCGCGAGATCCGTGGCCT	1920
Db	1861	CTATPCTGTGGGATGATCCTTTTAGAGAGAGGAGAGACTCCGCGAGATCCGTGGCCT	1920
Qy	1921	TATGTTTCCAGAGACTCTGCGCTCTTTCTATATGGAGAGTGTAGCTACGCGCATACGAT	1980
Db	1921	TATGTTTCCAGAGACTCTGCGCTCTTTCTATATGGAGAGTGTAGCTACGCGCATACGAT	1980
Qy	1981	CATGCGATGAAGACCGAGATCTACACCCGCCGCCCGCGACGCTCTTCGACGATCATACT	2040
Db	1981	CATGCGATGAAGACCGAGATCTACACCCGCCGCCCGCGACGCTCTTCGACGATCATACT	2040
Qy	2041	TCTTGGGAGAGATATGTCCTGGGCTGAGAGACTGGGAATCGAGTGTCTGTTGAAAATACC	2100
Db	2041	TCTTGGGAGAGATATGTCCTGGGCTGAGAGACTGGGAATCGAGTGTCTGTTGAAAATACC	2100
Qy	2101	AGCGGACAGAGATTTTTCGAGATACATCATCTATTTTAAAGCCCAAGCTGTTTACTCG	2160
Db	2101	AGCGGACAGAGATTTTTCGAGATACATCATCTATTTTAAAGCCCAAGCTGTTTACTCG	2160
Qy	2161	CGCCAAAGATACCTTTGTTGAACCTAGAGACTATCAGTGTGATTTTGTGATTCGCATCTT	2220
Db	2161	CGCCAAAGATACCTTTGTTGAACCTAGAGACTATCAGTGTGATTTTGTGATTCGCATCTT	2220
Qy	2221	TATTAACCTTGGGATTCCTCTTGGAAATCAGTTTAGAAGAACGTTTGCAGACAAATTTAT	2280
Db	2221	TATTAACCTTGGGATTCCTCTTGGAAATCAGTTTAGAAGAACGTTTGCAGACAAATTTAT	2280



9	341	13.5	10757	1	AE001628	AE001628 Chlamydia
C	110	341	13.5	12676	1	AE002192
10	341	13.5	12780	2	CPN133034	AE002192 Chlamydia
12	341	13.5	300650	2	AP002546	AE002546 Chlamydia
13	340.2	13.5	16448	1	AE001587	AE001587 Chlamydia
14	338.6	13.4	2787	9	AE001835	AE001835 Sequence 9
C	15	313.6	12.4	10757	1	AE001628
16	310	12.3	4926	2	CPU72499	AE001628 Chlamydia
17	309.4	12.2	2815	9	AE001829	AE001829 Chlamydia
18	309.4	12.2	12676	1	AE002192	AE002192 Chlamydia
C	19	309.4	12.2	12780	2	CPN133034
20	309.4	12.2	300650	2	AP002546	AE002546 Chlamydia
21	295.4	11.7	2757	9	AE001837	AE001837 Sequence 11
C	22	287.4	11.4	3000	9	AE001851
23	287.4	11.4	6030	2	CPOMP54	AE001851 Sequence 25
C	24	280	11.1	2793	9	AE001841
25	280	11.1	10026	1	AE002193	AE002193 Chlamydia
26	278.4	11.0	15068	1	AE001627	AE001627 Chlamydia
27	269.6	10.7	2787	9	AE001839	AE001839 Sequence 13
28	209.6	8.3	6110	2	CPU65942	AE001839 Chlamydia
29	209.6	8.3	6234	2	CPU65943	AE001839 Chlamydia
30	189.6	7.5	2838	9	AE002191	AE002191 Chlamydia
C	31	183.8	7.3	15789	1	AE002191
32	182.2	7.2	10695	1	AE001629	AE001629 Chlamydia
C	33	154.8	6.1	10044	1	AE002237
34	153.6	6.1	10766	1	AE001585	AE001585 Chlamydia
35	150.4	6.0	3052	9	AE001831	AE001831 Sequence 5
36	130.4	5.2	840	9	AE001843	AE001843 Sequence 17
37	99.6	3.9	1200	9	AE001853	AE001853 Sequence 27
38	81.8	3.2	1545	9	AE001845	AE001845 Sequence 19
39	77.8	3.1	14177	1	AE001360	AE001360 Chlamydia
40	74.4	2.9	14482	1	AE002293	AE002293 Chlamydia
41	62.2	2.5	216	2	CPU59306	AE002293 Chlamydia
42	57.4	2.3	455	2	CPU59307	AE002293 Chlamydia
43	52	2.1	1141	10	AX083744	AX083744 Sequence
C	44	49.8	2.0	1141	10	AX083744
C	45	49.2	1.9	5601	1	AE001361

## ALIGNMENTS

RESULT 1  
AB1833  
LOCUS AB1833  
DEFINITION Sequence 7 from Patent WO9858953.  
ACCESSION AB1833  
VERSION AB1833.1 GI:6731867  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
BASE COUNT  
ORIGIN

1 (bases 1 to 2526)  
MadSEN, A. and Birkelund, S.  
NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE  
Patent: WO 9858953-A 7 30-DEC-1998;  
MADSEN ANNA SOFIE (DK); BIRKELUND SVEND (DK)  
Location/Qualifiers  
1. 2526  
/db\_xref="taxon:32644"  
/organism="unidentified"

Query Match 100.0%; Score 2526; DB 9; Length 2526;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGAGATTCACCTCGCTTTTATGATATCATATGATACCTGCTTCTATGCGAAT 60  
Db 1 AAGAGATTCACCTCGCTTTTATGATATCATATGATACCTGCTTCTATGCGAAT 60

OY 61 TTATGAGAGCTGCTACTACCGAGAGCTATCGCTAGCAATAGCTTCGATGGAATAC 120  
Db 61 TTATGAGAGCTGCTACTACCGAGAGCTATCGCTAGCAATAGCTTCGATGGAATAC 120  
OY 121 TCAACACAAAGCTTTCTGATTAACATCATCGGCTACAGAGTGGCAACATTTGTTTT 180  
Db 121 TCAACACAAAGCTTTCTGATTAACATCATCGGCTACAGAGTGGCAACATTTGTTTT 180  
OY 181 AAGAGATTCGATGCTATGAGAAATGACCAAAACGGGAAACCTAGCTACTAGTGT 240  
Db 181 AAGAGATTCGATGCTATGAGAAATGACCAAAACGGGAAACCTAGCTACTAGTGT 240  
OY 241 TTTAAATATGACGCTGACGCTGAGATCTAAATTTCTGAGAGGGGATTTCTTACA 300  
Db 241 TTTAAATATGACGCTGACGCTGAGATCTAAATTTCTGAGAGGGGATTTCTTACA 300  
OY 301 TTTAGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
Db 301 TTTAGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
OY 361 AAGAGATTCACCTGCTATGAGAAATGACCAAAACGGGAAACCTAGCTACTAGTGT 420  
Db 361 AAGAGATTCACCTGCTATGAGAAATGACCAAAACGGGAAACCTAGCTACTAGTGT 420  
OY 421 GTGACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
Db 421 GTGACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
OY 481 AAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
Db 481 AAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
OY 541 TCCCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
Db 541 TCCCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
OY 601 GCGGAGAGATTCATACCAAAACCTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
Db 601 GCGGAGAGATTCATACCAAAACCTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
OY 661 GGAATATGAG 720  
Db 661 GGAATATGAG 720  
OY 721 ACCCTATTCATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
Db 721 ACCCTATTCATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
OY 781 ACAG 840  
Db 781 ACAG 840  
OY 841 CGGCTGCGGAG 900  
Db 841 CGGCTGCGGAG 900  
OY 901 TCTGTTGCTGATGCTGCTGATATATATAGCCCTGATATGAGAGAGAGAGAGAGAG 960  
Db 901 TCTGTTGCTGATGCTGCTGATATATATAGCCCTGATATGAGAGAGAGAGAGAGAG 960  
OY 961 GGAACCATAGCTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Db 961 GGAACCATAGCTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
OY 1021 CGGCTGCTGATGCTGCTGATATATATAGCCCTGATATGAGAGAGAGAGAGAGAG 1080  
Db 1021 CGGCTGCTGATGCTGCTGATATATATAGCCCTGATATGAGAGAGAGAGAGAGAG 1080  
OY 1081 GATGCTGTTTAAAG 1140  
Db 1081 GATGCTGTTTAAAG 1140  
OY 1141 TTAGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
Db 1141 TTAGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 00:58:17 ; Search time 18674.8 seconds

(without alignments)  
2092.202 Million cell updates/sec

Title: US-09-446-677B-7

Sequence: 1 ATGAGATTCACATCGCTT.....GTAGCAAAATCAATTTTAG 2526

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

GenDbml:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

```

```

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_rod:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_vil2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rod1:*
95: gb_rod2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2526	100.0	2526	9	A81833 Sequence 7
2	2510	99.4	26920	2	CPN133035 A1133035 Chlamydia
3	2508.4	99.3	10574	1	AE001586 Chlamydia
4	2508.4	99.3	12127	1	AE002235 Chlamydia
5	2493.2	98.7	299650	2	AP002545 Chlamydia
6	1425.2	56.4	1830	2	A81855 Sequence 29
7	341	13.5	3200	9	A81827 Sequence 1
8	341	13.5	6030	2	CPOMP54 A2001311 Chlamydia

QY 658 VNNETPKHYTSMASFQSLFSDKDYAVNNEXMYGLSYLYOYTSIGNIFRYASRNPV 717  
 Db 585 VYAOITSEDFVSAFQOLGKDKDYLVSKNSTVAGSIYOH-ISTYMTW----- 634  
 QY 718 MWGILSRFLQNP-----MIEFHCAYGHATNDKMTD---YA---NEPVMKSNMRN 764  
 Db 635 -----NTLLQNTLGAAPLVNAQALAYCHASNMTNTMDTAYAPKTTYSEIKGDMND 688  
 QY 765 CMAIEGGSMPLLVFENGRLFOGAIPEPKIQLYAYVQDKEFTAD-GRRFSGSLTSTS 823  
 Db 689 CFCVEEGANP--IETASLLEDMYSPFKQLVHAHQDDEKENNSDQGRYFESNNLTNLS 746  
 QY 824 VPLGIFPEKALASODVLYDFSEFSYIPDIFPKDSCCEALVYSGDS--WLYPAHVSRAH 881  
 Db 747 MPGVLEKFSHMDYASYNLTALAYADIVKSNPDCTASLIVSTSAVWYTKANNLRHAF 806  
 QY 882 VGSGTGRYHNDYTELLCRGSIIECRPHARNYNINCSKFRF 922  
 Db 807 ILQAGNYVALTRNTELFSGFELRGSCRYNIDLSKIOF 847

RESULT 15  
 Q9RB71 PRELIMINARY: PRT: 772 AA.  
 ID 09RB71:  
 AC 09RB71:  
 DT 01-MAY-2000 (TREMREL. 13, Created)  
 DT 01-OCT-2000 (TREMREL. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMREL. 16, Last annotation update)  
 DE PMP\_3\_2.  
 GN Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OC NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE=2030349; PubMed=10871362;  
 RA Shitai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 Shiba T., Ishi K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CWL029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL; AP002545; BAA98226.1; -  
 DR InterPro: IPR003357; -  
 DR InterPro: IPR003368; -  
 DR pfam: PF02385; OMP.1.  
 DR pfam: PF02415; DUF145; 1.  
 SQ SEQUENCE 772 AA; 82931 MW; 484FC56D635801EB CRC64;

Query Match 25.9%; Score 1249; DB 2; Length 772;  
 Best Local Similarity 37.3%; Pred. No. 2e-73;  
 Matches 296; Conservative 135; Mismatches 312; Indels 50; Gaps 20;

QY 157 FEQNOCKTKGGAISGANVTIGNYDSVSEFYQNAAT-FGGAIHSSGPIQIAVNOAEIRFAQ 215  
 Db 3 FSKNFSTONGGATKATLTLUTGTMSALFSEMSKKGAIQTSALJLTGNGEVSFSD 62  
 QY 216 NTKAKNGSGALISDGDIDDONAYVLFRENE-----ALTTAIGKGAVCCLPTSGSSTPY 270  
 Db 63 NMSSD--SGAIFTEASVTISNNAKVSFIDNKVYGASSSTTGDMSGAICAYKTSTDTK-- 119  
 QY 271 PIYTFSDNKOLVERHNSIMGGAIYARKLTISSGGPTLFINNISYANSOONLGGATAIDR 330  
 Db 120 --VTLGNOMLFLSNNTSTTAGCAIVKLELASGGLTLFSSNSVNGTAPKGGATAIED 177  
 QY 331 GGEISLSAEKGTTFQGN--RTSLPFLN--GIHLONAKFLKLOARNGCSIEFYDPT-- 384  
 Db 178 SGEISLSADSGDIVLFGNVTSTPTGTNRSIDLGTSAKMTALRSAAGRAIYFDPTTG 237  
 QY 385 SEADGSTOLININDPKNK--EYTGTLFSGEK---SLANDPRDFKSTIPQNVLSAGLYV 439

Db 238 SSTVTVDLVKVNETPADSALOYTGNIIFTEGKLESETEAADSKMLTSLKLOPYTLISGFTLS 297  
 QY 440 IREGAETYSKFTQSGSHLYLDLGTKLASKEDIAITGLAIDIDLSSTTAAYIKANT 499  
 Db 298 LKHGVTLOQAFQOADSRIEMDVGTL--EPADTSTINLVINISSIDGKKAR-LETRA 355  
 QY 500 ANKOISVTPSIELISPTGNAYEDLRMRNSOTFPLLSLEPCAGSVTVTADFLPVPSP--- 556  
 Db 356 TSKNLTSLGSTITLLDPTGTFTYENHSLRNQSYDILBK--ASGTIVSTA-----VTPDPI 408  
 QY 557 ----HYFGQNN-KIATGTG--NKVGEFFWDKINKPREKEGNLVNLMGNAVNRS 609  
 Db 409 MCEKFFHYGYOGTWGPYV--GTGASTATFETWTKGYIPNDERIGSLVPSNLMNAFTIDIS 467  
 QY 610 LMOVOTHASLQTDGGLMTIDGIGNFEHVSASDNIIRYRNSGCVLYSVNNETPKHYTS 669  
 Db 468 LHYMETANEGLOGDRAFWAGLSNFFHKDSTYTRRGFRSLSGYVIGGLHTCSKILS 527  
 QY 670 MAFSOLFSDKDYAVANNEXMYGLSYLYOYTSIGNIFRYASRNPVNVGILSRRELQV 729  
 Db 528 AAFCOLFGRDROVFAKNGQTVYGGTLYYOHNET-----YISLPCKLRCSLSYVPEL 581  
 QY 730 PLMIEFHCAYGHATNDKTDYANPPMVKNSMRNCMAIECGSMPLLVFENGRLFOGAI 789  
 Db 582 P-VLFSGNLSYHTDDLDLTKYTYTPVYKSGWGNDSFALEFGGRADICDESA-LFEQY 639  
 QY 790 PEKQLQLYAYVQDKEFTADGRRFSGSLTSTVPLGTRFEKALASODVLYDFSEFSYIP 849  
 Db 640 PEKQLQLYAYVQDKEFTADGRRFSGSLTSTVPLGTRFEKALASODVLYDFSEFSYIP 849  
 QY 850 DIFRKPSCCEALVYSGDSWLYPAHVSRAHVSCTGRYHNDYTELLCRGSIIECRPH 909  
 Db 700 DLYRSNPDCTTLIRISGDSKFTGTMLARQALVLRAGNHCFNSNEARISQFELRGSS 759  
 QY 910 RNTNINCSKFRF 922  
 Db 760 RNYNVDLGAKYQF 772

Search completed: October 2, 2001, 03:31:13  
 Job time: 253 sec

```

Db 35 LAFSCHLOSLANEEL-----LSPDSEFNG-----NIDSGTFPTKTSATYTLTGD 79
QY 62 VSISNV-DNSALNACFENVTSQVTFAGNHGILFNNISSTGTEGAVLCCODPQATARF 120
Db 80 VFEYEPKGTLSOSCFQOTTDNLFLNGHSLTFEGFIDAGTHAGAASTANKMLT--F 137
QY 121 SGFTLSFIOSPGD--IKEOGCLYSKNAFMLNNVVFREONOSKTKGASIGAVNTYVG 178
Db 138 SGFTLSFIOSSPTVTYTGQGTLSAGVNLNTRKLVVAGNFSTADGALKGASFLITG 197
QY 179 NYDSVFYQN-AAFFGGAHSSGFLQAVNOAEIRFAONTAKNSGGALYSDGDDIDON 237
Db 198 TSGDALFENNSSSTKGAIATTAGARIANNNGVYFLSNIAST--SGAIDDEGTIILNN 256
QY 238 AAYLFREHEALTTAIGKGAACCLPTSSSTPVPYVTSOMKQVLEFNHSHMGGAITYA 297
Db 257 KFLFEGNAATTT---GGAICNTRKASGS---PELLISNNKTLTLPASNAVETSGAIIHA 308
QY 298 RKLISISSGPTLFT--NNISYANSONMAGAIAIDPGEISLSAEKGTTFQGN-----RT 350
Db 309 KKLALSSGCFEFLRNNTSSATPK--GGAISIDASGELSLAEIGNITFVNTLTITGST 366
QY 351 SLPLNGIHLQNAKFLQARNGCSIEFYDPTISEADGSTQNLNIDPKN--KEYGTI 408
Db 367 DTPKRNAINISNGKFTELRAKKNHTIFFYDPTISEGTSVDLKINNAGALNPFQGTI 426
QY 409 LFSGEKSLANDPR--DEKSTIPQNVNLSAGVLYIKEGAETVSKFQSGSHLYLDIGT 465
Db 427 LFSGETLTADLKAADNKKSSFTQVSLSGKLLQKQVLTLESTSEFSDASLSLMDSGT 486
QY 466 KLIASKEDIAITGLAIDISLSSSTAIVIKANTANKOISVTDISELISPTGNAYEDIRM 525
Db 487 TLTSTAGSITTTNGLINVDISGLKQPVSLTAKGASNKI--VSGKMLNIDIGNIYESIMEF 545
QY 526 RNSGTFFPLSLTEPGAGSVYTAGDELFPV-----SPHYEGQNMKLAN--TGKNKVE 577
Db 546 SHDULFSILKITTVDADVTNDISLIPPAEDPNSEYEGFOGOMNVNMTDTATWTKKAT 605
QY 578 FFMKINKKPRKEGNNVPIINMGNAVNRSIMQVOETHASSLOTGMLIDGIGNEFH 637
Db 606 ATWKRTGFPVPERKSALVCTLMGVTDIRSLQDLVIGATGKMHKGGFVSSSTNPLH 665
QY 638 VASSEDNIRYRHNSSGYVLSVNNELTPK--HYTSMASFOLFESRDQVAAVSNNEYRYMAGSY 696
Db 666 KTGENRKGFEHHTSGGYVIG--GSAHTPRKDLFTFAFCHLFAKDCKCFIAHNNSRYGCTL 724
QY 697 LYQTTTSGNIFRASRPNVNVGILSRFLQNL-----MIFHFLCAYGAHTDNMK 748
Db 725 FFKHSHTL-----OPONYLRLGRAKFFESAIEKPREPIPLALDVQVSFSDHDFME 775
QY 749 TDVANFPMVAKSWNNKCAITEGCSMPPLVPENGRLFGCAIPKMLQIVVAYOGDFKETT 808
Db 776 THYISLPSSEBSWSNECLAGIGLDLPVLSNPHRLFTFTIPQKAVEVNYVSQNSFFESS 835
QY 809 ADGRRFNSGLTSSVPLGIRFEKRLASQDVLVDYFSESYIDIRFKDPSCEAVALISGDS 868
Db 836 SDGSGFSIGRLNLSIPVAKAFVQGDIGDSDTYDLSGFEVSDVYRNNQSATLVMSPDS 895
QY 869 WLVAHAHVRARAFVSGTGRHENDYTELLRGSTIECPHARRNTININGCSFFR 922
Db 896 WKIRGGLNRQAFILRGSNVYVNSNCELFGHYAMELRGLGSSRNYNVVDGTKLRF 949

```

```

OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RX Longline-96406378; PubMed-8810511;
RA Longbottom D., Russell M., Jones G.E., Latson F.A., Herring A.J.;
RT "Identification of a multigenic family coding for the 90 kDa proteins
  of the ovine abortion subtype of Chlamydia psittaci.";
RL FEMS Microbiol. Lett. 142:277-281(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RX MEDLINE-98187897; PubMed-9529048;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RT "Molecular cloning and characterization of the genes coding for the
  highly immunogenic cluster of 90-kilodalton envelope proteins from the
  Chlamydia psittaci subtype that causes abortion in sheep.";
RL Infect. Immun. 66:1317-1324(1998).
DR EMBL; U65942; AAC15921.1; -
DR InterPro; IPR003357; -
DR InterPro; IPR003368; -
DR Pfam; PF02385; OMP; 1.
DR Pfam; PF02415; DUF145; 1.
DR Pfam; PF02415; DUF145; 1.
SQ SEQUENCE 847 AA; 90694 MW; 754C938E7F11179E CRC64;

Query Match 26.1%; Score 1258; DB 2; Length 847;
Best Local Similarity 33.0%; Pred. No. 6.1e-74;
Matches 311; Conservative 150; Mismatches 310; Indels 170; Gaps 28;

QY 30 SLTEPDSFHSQONAE-----RSYVQADGVYSLNGVSI--SNVNSALNACFNVTSGS 83
Db 29 TLTASHVSNQMT--NSEPNNPLSTNSNGTITTCGNICIAVAGDGLSSSCFFDTAGN 87
QY 84 VTFAENHGLYFNINISCTTEGAVLCCODQAPARRSGESTLSFIOSPGDIKEGCLYS 143
Db 88 LSLFNGCYTLCFDNTTQSSHPGAI--SVSGTNKTLIDISGFLFS-----CAVC 134
QY 144 KNALMLNNVYVRFQONSKTKGASISGANVTIVGNDSVFQONATFGGAHSSGFLQ 203
Db 135 P-----PGATGYGAIKAVGNTTIDNSSLVFNKCS-----GGAIAQCKAS 166
QY 204 IAVNOAEIRFAONTAKNSGGALYSDGDDIDONAAVYLFREHEALTTAIGKGAACCLPT 263
Db 167 -----GGAIAQCKAS 177
QY 264 SCSSTPVPYVTSOMKQVLEFNHSHMGGAITYARKLSSSGPTLFINNISYANSONLG 323
Db 178 SSEAE---LKIENNQNVFAENSSSSSGAIVADKLTVISGGPTLFSNN--SVASASSPK 232
QY 324 GAIAT--DTGGEISLSAEKGTTFQGNR-----TSLPFLNGIHLQNAKFLKQARNGC 375
Db 223 GATCIKDSGECSLFADIGDITFDGNKIITKNGSSPYTRNSIDLGSGKTKLNAKEGF 292
QY 376 SIEFYDPTISEADGSTQNLNIDPKNKEYTGTILFSGE-----KSLANDPRDFKSTIP 428
Db 293 GIFYFDPT--GGSDDELININ--KQDVTDYTKIYFSGERLSDDEKRVANLKSDFK---- 345
QY 429 QNVNLSAGVLYIKEGAETVSKFQSGSHLYLDIGTKL---TASKEIDATGLAIDIS 485
Db 346 QPLKISGSLILKDKGVLETFSQTEGATVVMQDGLTTLQPPSSGGEFTITLTINDIVAS 405
QY 486 LSSSSTA---AVIKANTANKOISVTDISELISPTGNAYEDLRMNSQTFPLLSLEPGAG 542
Db 406 LGGGGVAPDPAKVEATKESKVTI--NAVNLVDNGMALEYPIILASQPTAIEVRSSGSG 464
QY 543 SVYTAGDPLFVSP--HYFGQNMKLAN--TGKNKNGEFPKINYPKPREKGNLVP 597
Db 465 SITPPTNLENTPTPHYGYGNMTVTKGSSSAQEKATITLTWQGTGSPREPGQSLVP 524
QY 598 NILMGNAVNRSIMQVOETHASSLOTGMLIDGIGNEFHVSASEDNIRYRHNSSGYVLS 657
Db 525 NTLNGSPSDIAIQLNMDISVNGADYHRGFWVSGLGNFLHKSQSDTKRKRFHNSAGVALG 584

```

QY	1539	CTGCGATACAGGCAAAATGCCATATGACATGCTCCGAAGATTAATTCACAGAGCTTCCCTCT	1538
Db	1575	ggaaccgagcgggacggtttatgaaatacatgctttaagaatactccacgctccctacgacat	1634
QY	1599	GCCTCTCTTT--AGAGCTGAGACCCGGGGGTAGTGTGATCTGTACCTGCTGGAGATTCCT	1655
Db	1635	cttagagctcaaaagctcttcggaactctaaacaagcaaccgagagctccagatccatataat	1694
QY	1656	ACCGGTAAGTCCCATTAATGCTTTCAAGGCAATTGG--AAATTAGCTTGACAGGACAC	1712
Db	1695	gggtgtaagaaatccatcaatcagcgtcatcaaggaaacttgggcccgaattgcttggggagcag	1754
QY	1713	TGCAAAACAAGTTGGAGAAATCTCTGGGATTTAAATTAATTTAAAGCCCTAGACCGTA	1772
Db	1755	ggctcttaagactcycaacctctcaactcagacatgaactctatctatctcctaaccgagcg	1814
QY	1773	AGAAAGAAATTTAGTTAGTTATGCTTAATATCTTGTTGGGGGAATGCTGTAATGTCAGATCCTTAAT	1832
Db	1815	tatcgctctcttagtcccatatagactctatgagatgcatattatagattatagctctctcca	1874
QY	1833	GCAGGCTCAGAAACCCATGCAATCCGAGCTTACAGACATGACAGGTGAGGGCTGTGTGATGATGG	1892
Db	1875	ttaccttaatgagacatcgcaaacgaagggtctgagggagacgctgtcttcttgytgcgcg	1934
QY	1893	AATTGGAAATTTCTTCATGATATCTCCCTCCAGACAAATATAAGTACCTGATTAACAG	1952
Db	1935	attatcaactctctccataaagatagatcaaaaaacagcgcggttctgcgaattgag	1994
QY	1953	CGGTGATATGTTCTATCTATGTAATATATAGATTCACACCTTAAGCCTTACTTGATGATGC	2012
Db	1995	tgggggttatgcatatagaggaaacctataactctgttcgaataagattcttagtctgc	2054
QY	2013	ATTTTCCCACTCTTTAGTAGACACAGACATATCGGGTTTCCACAACAGAAATACAGAT	2072
Db	2055	atttgcgcagctctcttgggaagagatagagactcttggagtaagaatacaaggtacgct	2114
QY	2073	GTATTTAGATCGTATCTCTATATATATACAACTCCCTAGGGAAATATTTCCGTTATGC	2132
Db	2115	ctaaaggaggaactctctatataccagacaagaacac-----tat	2154
QY	2133	TTGCGCAACCCCTAATGTAACGTGCGGATTCCTCCCAAGAAAGTTCTTCCAAATCCCTCT	2192
Db	2155	atctctctctcttcgcaactaagaagccttgcgttgcgtcttaagcttccacagagattccct	2214
QY	2193	TATGATTTTTCATTTTGTGTGCTTATGGTCATGCGCAACATGATATGAAGAACAGACTA	2252
Db	2215	gtctctcttcaggaacctgac-tacaacatcaggttaacgctctgaacaaccaagta	2273
QY	2253	CGCAAAATTTCCCTATGCTGTAAGAAACAGCTGAGAAACAATTTGTGGCTATATAGATGCGG	2312
Db	2274	tacaacatctctacatgcttaagaagacgtgggggaatgataagcttgcgtcttlaagaaltcg	2333
QY	2313	AGGAGATCCCTCTATTTGGTATTTATGAGAAGCAAGACATTTTCCAAAGTGCACCCCAT	2372
Db	2334	tgggaagagcccgcatcttgcttgaatgaaagtgc---ctatttggcagatataagccctt	2390
QY	2373	TATGAACATACAAATTAAGTTATGCTTATCAGGAGAAATTTCAAAGAGACGATCGATGATG	2432
Db	2391	catbaaattcgaattgctctatgacatcaaggaagttttaagaacacaggaaacaagac	2450
QY	2433	CCGATAGTTTATGATATGAGGAGTTTAAACATCGATTTTCTGACTCTCAAGGCAATACGCTTTGA	2492
Db	2451	tcgtgaatttgggaatgacgcgctctgtgtaactcttgcctcaactacatcgggataccgatttga	2510
QY	2493	GAAGCTGCACTTTCACAGATGTACTATATGACTTTAGTTTCTCCCTATATTCCTGATAT	2552
Db	2511	taagaaatcagactgcgaagatgcaacgtacaactaactctgtgtatactctgagatct	2570
QY	2553	TTTTCCTTAAGGATTCCTCATATGTAAGCTCTCTGTGGTATATAGCGAGACTCTGGCTTGT	2612
Db	2571	tgtctctagtaaacaccgacatgtaacgaacaactcgaaatttttgcgttgcgttcttcttggaaac	2630

Oy		2613	TCCGCGACACAGCATATCAACAATGCCTTTTGTAGGGAGTGGAACGGGTGGTCATT	2672
Dd		2631	cctgcgtacgaattgcaagacaagaactttagtcctcgtcgcaagggaaccaatttgct	2690
Oy		2673	TAAAGACTATTAAGTAGCCTTTATGTGCGAAGAATAGAAATGCCGCCCAATGCTAGAA	2732
Dd		2691	tactccaattttgaagacctttagccaattttcttcttgtaattgsgtgcatactgcaa	2750
Oy		2733	TTATATATATTAACATGTGTGGAGCAAATTTTCGTTTTTA	2768
Dd		2751	ttaccatgtagacttagagcaaaataccaatctcta	2786
 RESULT 7 AAAA48838 ID AAA48838 standard; DNA; 3050 BP. XX AC AAA48838; DT 15-SEP-2000 (first entry) XX Chlamydia pneumoniae 98 kDa outer membrane protein CPN100640 gene. DE Chlamydia pneumoniae 98 kDa outer membrane protein; antigen; immunogen; infection, KW vaccine; antibacterial; community acquired pneumonia; bronchitis; KW sinusitis; acute respiratory disease; upper respiratory tract disease; KW asthma; atherosclerosis; ds. XX OS Chlamydia pneumoniae. XX FH Key Location/Qualifiers FT CDS 101..2911 FT /*tag= a FT /product= "98 kDa outer membrane protein" FT sig_peptide 101..133 FT /*tag= b FT mat_peptide 134..2911 FT /*tag= c XX PN WO20032784-A1. PD 08-JUN-2000. PF 01-DEC-1999; 99WO-CA01148. PR 01-DEC-1998; 98US-0110439. PR 03-MAY-1999; 99US-0132272. XX (CONN-) CONNAUGHT LAB LTD. PA Murdin AD, Oomen RP, Wang J, Dunn P; XX WPI: 2000-412330/35. DR P-PSDB; AAY99842, AAY99843. XX PT New polynucleotide encoding the Chlamydia 98 kilodalton outer membrane PT protein, useful for preventing or treating Chlamydia Infection - XX PS Claim 2; Fig 1; 98pp; English.				
XX	The present sequence encodes the Chlamydia pneumoniae 98 kDa			
CC	outer membrane protein, known as CPN100640. Chlamydia pneumoniae			
CC	is a common cause of community acquired pneumonia and upper			
CC	respiratory tract symptoms and diseases, including bronchitis and			
CC	sinusitis. It also has an association with atherosclerosis and asthma.			
CC	The 98 kDa outer membrane protein is a C. pneumoniae-specific antigen			
CC	which can confer immune protection against chlamydial infection. The			
CC	nucleic acid sequence encoding the protein or the protein itself may be			
CC	administered as a vaccine to prevent or treat infection and they may			
CC	also be used to diagnose infection. The gene encoding CPN100640 was			
CC	amplified from Chlamydia pneumoniae genomic DNA by PCR.			
XQ	Sequence 3050 BP; 850 A; 735 C; 604 G; 861 T; 0 other;			

Query Match	8.5%;	Score 260.8;	DB 21;	Length 3050;
Best Local Similarity	48.8%;	Pred. No. 3.6e-64;		
Matches 1055; Conservative	0;	Mismatches 1032;	Indels 75;	Gaps 10;

QY	657	GAATGGTTGGGAGGGGCTTTGGTACGCCGATGGTGATATTGATATTGATATCGAATGCTTA	716
Db	778	gaaagagctctgggcggcgctctctgacagcactgacacactcttcgatatacaagaattcttca	837
QY	717	TGTTCTATTTTGGAGAAATGAGGACATTTGACTACTCGCTATAGGTTAAGGAGAGGGCTGTGCG	776
Db	838	agtgatctttgacgccaatgctgtcttgggaagccgtctcaagctcaggcggtctatttg	897
QY	777	TTTGCTTCCCACTTCAGGAAGTAGTAGTACCTGACGTCTCTATTGTGACATTTCCTGGACAAATA	836
Db	898	ttgc-----actacgacataaanaaagttacttacttactggaagaa	939
QY	837	ACAATTAGTCTTTGAAAGAAACCATTCCATTAATGGGTGGGCGGAGCCATTATATGTCAGGAA	896
Db	940	aaacctctttcacaaataatacagcattgcaatgtgcgagccatctctgacatca	999
QY	897	ACTTAGCATCTCTTCAGAGAGGCTCTACTCTATTATATCAATAATATATCATATGCAATTC	956
Db	1000	gltcagattcttcgcgtggagttcctactctacttccaagtaatactctcagaagtaagtc	1059
QY	957	---GCAAAATTTAGTGGAGACCTATTCCCATTTACTGAGGGGAGATCAGTTTTCAGC	1013
Db	1060	cggtcagggggggggggggcgatcaatataagcatctgtgtgggaacctctctctgc	1119
QY	1014	AGAGAAAGCAACATTACATT-----CCAAGAAACCGGACGACCTTACCGCTTTT	1064
Db	1120	tactctcggagatatttacttccaataacaacaaagtcacaaaggaaagcacaaagtaaca	1179
QY	1065	GAATGGCATCATCTTTTACAAATATCTTAAATTTCCGTAATATACAGGAGGAAATGATG	1124
Db	1180	aaagcacaataatatactatgtatcgcgtaaagtcacatcgtatacgaagctgtctaagggca	1239
QY	1125	CTCTATTGAATTTTATGATCTTATTACTTCTGAGACAGATGGGCTTACCC-----	1174
Db	1240	atctatctatcttatactatgataccctcacaaacccaggaaacggagcttcttaccgacat	1299
QY	1175	--AATTGAATATACAGGAGATCTTAAATATTAAGAGTACACAGGAGACCATACTCTTTTC	1232
Db	1300	gaacttaacttagcagatgbcgaaacagtgagatcagatctgggggtgagattgtcttcttc	1359
QY	1233	TGGGAAAAAGTCTAGCAAAACA-----TCTTAGGATTTTAAATCTACAACTCC	1283
Db	1360	tgggaaaaagcttccctctacagaaaaagcaatcgcgtgcaaacgctcacctctacatcgc	1419
QY	1284	TCAAAACGTCAACCTGTCGACAGATACCTTAAGTTAATAAGGCGGGCGGAATGCACAGT	1343
Db	1420	acaacctgcagattagcggcggggagatctgtactcgtgtgtgagtcagtcacgtaactt	1479
QY	1344	TTCAAAATTACGCGACTGTCACAGATTCGCATTAGTTTATAGATTAGGACCAACATGAT	1403
Db	1480	caagatcttactcaagaatccagatcccgcaacttaatgtagtggggagactaaccttag	1539
QY	1404	AGCCTCTAAGGAGACATTGGCCATCCACAGCCCTCGGAGATAGATTAGATGCTTAAGCTC	1463
Db	1540	tgtctaaaggagcaaatcttgcgtctaattgctcttagcagtaaatctctcctcttgaatag	1599
QY	1464	ATTCCTCAACGACGCTGTATTATTAAGCAAAACCGCAATTAACAGATATCCGTGACGA	1533
Db	1600	aaccacaagaagcagctt---aaaaacagaagctgcaagataaataatcagcctatcg	1656
QY	1524	CTCTATTGAACCTTATGTCGCTACTGCTACGAAATGCTATGAAATGTCACAAATGAGAAATTC	1583
Db	1657	aacgatgtgcctatctatgacacggaaaggtcattctatgagatcaacttaactaaaaatg	1716
QY	1584	ACAAGCTTCCTCTGCT-----CTCTTTAGACCTGGAGCGGGGGGTAGTGT	1631
Db	1717	tagtactaactacccctctcttgaacttaaccacgcgaaagacgaacaaagatattactctgg	1776

QY	1632	GACGTACACTGCTGGACATTTTCTACCGGTAATCCTCCCATTTATGTTTCAAGGCAATG	1691
Db	1777	agctcttctctaacctgacctcttcgaagaacctgaaacccactacgggtlctcaaggaaactg	1836
QY	1692	GAATATTGACTTGGACGA---GGAACTGGAAAACAAGTTGGAGAAATTTCTTGCGATMAAT	1748
Db	1837	gcagttgtcttggcgcaaahtgcacatccctcaaaaataagagcatcaatctggaacccgtac	1896
QY	1749	AAATTTATAGCCTTAGACCTGAAAAAGAGAAATTTAGTTCTTAATATCTTGTGGGGAA	1808
Db	1897	aggaataccttctagcttccctgagaaaaaataatctccctctcaaatagcttcttggagaa	1956
QY	1809	TGCGTMAATCTGAGATCCCTTAATGCAGGTTCAAGGACCCCTGCATCGAGCTTACACAC	1868
Db	1957	cttataataatacgtctcgatcaatcgaacttgaacttataagaacaagtcagtgggagcctt	2016
QY	1869	AGATCGAGGGCTGTGATTCGATGGAATTTGGGAATTTCTTCATGTATCTGCTCCGAGA	1928
Db	2017	tgagcgttgagctatggtcttcgaagtcttgaaattcttctctataagattctatgcccac	2076
QY	1929	CAATATAGGTACCGGTATTAACGCGGTGATATGTTCTATCTGTAAATATAGAGATCAC	1988
Db	2077	ccgcgaatggttccgcgcatactacgcgggggtlctatgactaagatcatcacgaacaactcc	2136
QY	1989	ACCTAAGCACTATCTTCGATNGCATTTTCCCAACTTTTATGAGACAAAGACTATGC	2048
Db	2137	tgccgagatcatgactcttcttgcctcttgcgcagccttcttgcctagatagtcgaatcatat	2196
QY	2049	GGTTTCCACAACGAATATACGAATGTATTTAGAGTGTATCTATCATTAATACAACTC	2108
Db	2197	tacagtaagaacacagagataactaagctgctcttctgtatcttccacacatacagaag	2256
QY	2109	CGTAGGGAATTTTTCGTTATGCTTCGCGTAACCTTATGTAAAGTCGCGGATCTCTC	2168
Db	2257	gctcttcgaactcgcgaatttctctctgaggaaaaagcaaccgagctcccttgggtcttc	2316
QY	2169	AAGAAGGTTTCTTAAATATCCTCTATGATTTTTCATTTTGTGTGCTTATGTGTCATC	2228
Db	2317	tgagatctcccaagctcatctctctatctcg---ttcgatgtlaaatlcagtaatctccatac	2373
QY	2229	CACCAATGATATGAAAACACACATACACCAATTTCCCTATAGTGGAAAAACGCTGGAGAA	2288
Db	2374	agacaacaacaatgaaagacatactataacacgtatacctatcaacaagggtcttggagaa	2433
QY	2289	CAATGTGTGGGCTTAAAGTGCAGAGAGACATGCTCTATTTGATTTTGAAACGAGAG	2348
Db	2434	cgatgacctctctgacgatactcttggagctagccctgct---tttgtattctcgttcgta	2490
QY	2349	ACTTTTCCAGGTGCTCATTTCCATTTATGAACATCAATTAAGTTTATGCTTATCAGGAGA	2408
Db	2491	tctcttgaagaagatcgaaactcttctgcaagaatagctatactatgctgcatcagaaga	2550
QY	2409	TTTCAAGAGACGATCGATGATGCGGCTGATATTAATATGAGATTTAAACATCATTTTC	2468
Db	2551	cttctaagagcgctatgctctgaagaacgaccttcaataaagcggctatcaacgtata	2610
QY	2469	TGTACTCTTAGCGCATPACGCTTTGAGAGAGCTGCGACTTTTTCAGATGTACTCTATGACTT	2528
Db	2611	gattccctataagcgctcaactctgaaagagctcaaaatcagaanaaggaaacttaagact	2670
QY	2529	TAGTTTCTCCATATTTCCCTATATTTTTCGTAAGATTCCTCATATGAAAGCTGCTGTGT	2588
Db	2671	tactcttatgataatactgactgcttaccagaaacactcttaaaatgcaaatctccctaat	2730
QY	2589	GATTAGCGGAGATCCCGCTGTTTCCCGAGACACAGTATCAACATGCTTTTGTAGG	2648
Db	2731	agctagagatgctaacactgagtgctctatgtaaccaacctcgaaagaaagtttctcgt	2790
QY	2649	GAGTGGAAACGGGTGGGTATCATCTTTAAACGACTATACTGAAGCTTTATGTTCGAGCAATAT	2708
Db	2791	tcggtgtcgaaacatttccaagttaagtaaccacacatgagaaatctcttggtaalttgcctt	2850





```

QY 1749 AAATTATAGCTAGACCTGAAAAAGAGAAATTTAGTCTTAATATCTTGTGGGGAA 1808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109738 AGATACATCTCTAGCTGAGAGAAAAAGATCTCCCTTAATATAGCTTATGGGAAA 109679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1809 TGCCTAAATGTCAGATCCCTTAATGACGTTCAAGAGACCCTCATCGAGCTTACAGAC 1868
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109678 CTTTATGAAATATAGCTGATCATCATCTTAATAGAAACCAAGCCAGTGGAGCCCTT 109619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1869 AGATCGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1928
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109618 TGAGCGGAGCTATGCTTTCAGGAATTTGCAATTTCTCATAGACATCTATGCCAC 109559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1929 CAATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1988
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109558 CCGCATGATGTTCCGCTATGATGATGATGATGATGATGATGATGATGATGATG 109499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1989 ACCATACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2048
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109498 TGCCGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 109439
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2049 GGTTCACACACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109438 TACAGGTAAGACACGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 109379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2109 CTTAGGGAATATTTTCCGTTATGCTTTCGCTAACCTTAATGTAACGTCGGGATCTCTC 2168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109378 GCTCTTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 109319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2169 AAGAAAGTTTCTTCAAAATGCTTATGATGATGATGATGATGATGATGATGATGATGATG 2228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109318 TGAGATCTCCAGATCTTCTTATGCTGATGATGATGATGATGATGATGATGATGATGATG 109262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2229 CACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109261 AGACAACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 109202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2289 CAATGTTGGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109201 CGATGCTTCTGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 109145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2349 ACCTTCCAAAGGTCCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109144 TCTTCTAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 109085
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2409 TTTCAAGAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109084 CTTCTAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 109025
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2469 TGTACTCTTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2528
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109024 GATTCCTATAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 108965
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2529 TAGTTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108964 TACTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 108905
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2589 GATTAGCGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108904 AGTAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 108845
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2649 GATGAGGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108844 TCTGCTGAGCAACCATTTCCAAATGAAACCCCAATGGAATCTCGCTCAATTCGCTTT 108785
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2709 AGATGCGCGCGCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108784 TGAAGTACGAGATTTCTACGAAATTTATTAATACAACTAGGCTTAAGTTTGTTCCTA 108725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2769 GA 2770
    |||
Db 108724 GA 108723

```

```

RESULT 9
AAA30847
ID AAA30847 standard; DNA; 2950 BP.
XX
AC AAA30847;
XX
DT 29-AUG-2000 (first entry)
XX
DE Chlamydia antigen CPN100634 full length coding sequence.
XX
KW Chlamydia antigen; diagnosis; infection; community acquired pneumonia;
KM therapy; upper respiratory tract disease; bronchitis; sinusitis;
KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;
KM ds.
XX
OS Chlamydia pneumoniae.
XX
FH Key location/Qualifiers
FT CDS 101..2887
FT /tag= a
FT /product= Chlamydia antigen CPN100634
XX
PN NC0200032794-A2.
XX
PD 08-JUN-2000.
XX
PE 01-DEC-1999; 99MO-CA01147.
XX
PR 01-DEC-1998; 98US-0110339.
PR 01-DEC-1998; 98US-0110340.
PR 01-DEC-1998; 98US-0110427.
PR 01-DEC-1998; 98US-0110428.
PR 01-DEC-1998; 98US-0110438.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Mirdin AD, Oomen RP, Wang J;
XX
DR WPI: 2000-412339/35.
DR P-PSDB; AAY90236.
XX
PT Nucleic acids encoding polypeptide antigens from Chlamydia useful for
PT preventing, diagnosing and treating diseases such as community acquired
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
PT asthma -
XX
PS Claim 2; Fig 1; 174pp; English.
XX
CC This sequence encodes a Chlamydia antigen of the invention, designated
CC CPN100634. The nucleic acids (and their complementary sequences) may be
CC used as diagnostic agents for detecting the presence of nucleic acids
CC encoding Chlamydia antigens in samples according to standard methods,
CC and therefore, for diagnosing Chlamydia infections. For example, they may
CC be used as primers and probes for diagnostic polymerase chain reaction
CC (PCR) assays. Antisense sequences may be used to down regulate
CC expression of the proteins and may be used to treat infections. The
CC nucleic acids may also be used to produce the protein antigens they
CC encode according to standard recombinant DNA methodologies. The
CC proteins may then be used as antigens for the production of antibodies
CC (i.e. as vaccines) for preventing infection by Chlamydia. The
CC antibodies may also be used as diagnostic reagents for detecting
CC infections. Chlamydia is a pathogen implicated in the development of
CC (for example) community acquired pneumonia, upper respiratory tract
CC disease (especially bronchitis and sinusitis), asthmatic bronchitis,
CC adult-onset asthma and acute exacerbations of asthma in adults.
XX
SQ Sequence 2950 BP; 851 A; 670 C; 596 G; 833 T; 0 other;

```

Query Match 8.3%; Score 252.6; DB 21; Length 2950;  
 Best Local Similarity 46.5%; Pred. No. 7.8e-62;  
 Matches 1292; Conservative. 0; Mismatches 1389; Indels 96; Gaps 11;





QY	567	AGCCACATTTTGGAGGCTCTATTCATCTTTCAGGTGCCCTCCACAGATTTGCAGTAATTAACAGC	626
Db	768	atcaacaacaaaggaggaacaaattgctactacacagcagcgctgcctacagcaataaacacag	827
QY	627	AGAAATAGATTTGGACCAAAATACTCCCAAGATGGTTCTGGAGGGCGTTTGATCCGA	686
Db	828	ttatgttagttctctactctaacctagc---gtctacgtcagagcgctatcgtatgata	884
QY	687	TGCTGATTTGATTTGATTCAGATCCATCTTATGTCTTATTTGCAGAAAATGAGCATTTGAC	746
Db	885	aggaacgctcactactcgaacacaacaaattctataatttgaaggaatgacgcgnaaac	944
QY	747	TACTGCTATAGTAAGGAGGAGGGCTGTCTGTTGTCTCCACTTCAGAGAGTACTCC	806
Db	945	tactg-----gaggtcgcatctgcaacacaaaggcgatg	980
QY	807	AGTTCCTATTGTGACTTTCTCTGACATAATAACAGTTAGTCTTTGAAAGAAACCATTCAT	866
Db	981	atctccgaactgataatctacttaacataaagactcgtactctgtcttcacaaactagcaga	1040
QY	867	AATGGGTGGGGGCAATTATTCGTGTGGAAACCTTAGCATCTTCAGAGGGCTCCTACT	926
Db	1041	aacaagcggcggccacccacacagcctaaagaagccttcccttcggaggttatacga	1100
QY	927	ATTATATCAATATATATATCATATTCATAATTCGCAAAATTCGCAAAATTTAGGTGGAGCTTATTCGATTA	986
Db	1101	gttctctacgaataatagctctcactcagcaactcctaaag---gggggtgcctatcagatcga	1157
QY	987	TACTGGAGGGGAGATCAGTTTATCAGCAGAGAAAGAACATTTACATTTCCAAAGAAA---	1043
Db	1158	tgctccaggaagacctcagctctctctctgcagagacgaacatcttacttgaagaataac	1217
QY	1044	-----CCGGACGAGCTTACCGGTTTGAATGGCAATCCATCTTTTCAAAA	1088
Db	1218	ccttaacaacaacgggaagtccgatactctcttaacgtaatgysgatacaatacagaagtaa	1277
QY	1089	TGCTAAATTCCTGAAATTTTACAGGCGAGAAATGGATGCTCTATAGAAATTTATGATCTAT	1148
Db	1278	cgggaaatccaggaattacggcgtcgtctaaataatcatalcaattctctctatgctccat	1337
QY	1149	TACTCTGGAAGCAGATGGGCTCTACCCCAATTTGAAATATCAACGAGATNCCG-----AAAA	1202
Db	1338	caacttcagaagaacctcacaagacgtaattgaagaataaataacgctccgcggagctct	1397
QY	1203	TAAAGATACACAGGAGCCATCTCTTTCTGTGAGAAAAGACTCTAGCAAAAGATCTAG	1262
Db	1398	caatccatcatcagaagcgtctctattcttctcgagaaaccttaacagatgtaacttaa	1457
QY	1263	GG-----ATTTAAATCTACATCCCTCAGAAACGTCAACCTGTCTGCGAGATACTT	1313
Db	1458	agttgctgacaatttaaaatcttcatctacacgaacgacgtctccatccctcggaggaagt	1517
QY	1314	AGTTATTTAAAGAGGGGGCCGACATCTCTTTTCTGAGAAAAGACTCTAGCAAAAGATCTAG	1373
Db	1518	atgtctacaagaagggtcactctttagaaggcaagcgtctctctcaaggggcggtctct	1577
QY	1374	TTTATGTTTATGATTTAGAACCAAACTGATAGCCTCTTAAAGAAAGCAATTGCCATCACAG	1433
Db	1578	ccttcgcatgtatccaggaacgacatcttaacatacagcttgtagtatttaacatcagaa	1637
QY	1434	CCCTGGCATGATATATGATATGCTTTAGCTCATTCCTCAACAGCAGCTGTTTATTTAAAGCAA	1493
Db	1638	ccttagaatacaagtcgaacctcttaggtcttaagcagcccgctcagccctaaacagcaaaag	1697
QY	1494	CACCGCAATTAACAGATATCCCTGACGAGACTCTATAGAAATTTATCTCGCTACTAGCGAA	1553
Db	1698	tgctccaataaagtagt---cgtactcgggaagctcaaacctcgtatgtatattgaaggaa	1754
QY	1554	TGCTATGAAGATCTCAGAAATGAGAAATTCACAGACGTTTCCCTCTGCTCTCTTTAGAGCC	1613
Db	1755	catttaagaagcatatgcttcagccatgacagacgctctctctctatctaaataacaggt	1814

OY	1614	TGAGCCGGGGGTAAGTGTACCTTAACTGCTGAGATTTCCTACCGGT-----	1661
Db	1815	tgatgctgagatgtgatactaagcttgatacaacgagccttaacccctgtctcgtcgaaga	1874
OY	1662	---AAGTCCCATTAATGCTTTTCAAGCAATAGGAATTAATGCTTGG-----ACAGG	1709
Db	1875	tctcaatcagaatacaggaattccaaggaacaagaaagttaatttggactaagatacagc	1934
OY	1710	AACGTGAACAAAGTTGAGAAATTCCTTGGGATAAATAATTAATTAAGCTTACCTGA	1769
Db	1935	tacaatacaaaagagcgccagcgcaacttggaccaaaacaggaattgtctccagcccca	1994
OY	1770	AAAAGAAGAAATTTAGTCTCTAATATCTTTGGGGGAATGCTGTAAATGTCAAGTCTT	1829
Db	1995	aagaaaatctgcgttagtatacctaaccctabagagagcttctaagatcttgcctct	2054
OY	1830	AATCAGGTTCAAGAGACCCATCAGTACAGACCTTACAGACAGATCAAGAGGCTGTGATGA	1889
Db	2055	gcaacagctctgtagagatcgcgcaactcgttatgaaaccaacaagaattctcgggttc	2114
OY	1890	TGGAATTTGGGAATTTCTTCCATGTATCTGCTTCGGAAGCAATATTAAGTACCGTATTA	1949
Db	2115	ctccatgacgaactctcctgataaagcttgaagatgaaatcgcaaggtcttcgcataac	2174
OY	1950	CAGCGGTGATATTTCTTATCTGTAAATAATAGATACACCTTAAGCACTATCTGCAT	2009
Db	2175	ctcggagagctacgctcagctggtgaagtgctcacactcctaagaagcactataactct	2234
OY	2010	GGCATTTTCCCACTCTTTAGTAGAGACAAAGGACATAGCGGTTCCCAACAGAAATACG	2069
Db	2235	tggctctgcacatcctctgtcagaagaacaagttgttatacgcctcaacaacactcag	2294
OY	2070	AATGTATTTAGATCGATCTCATATCAATTAATACACCTCTAGGAAATATTTTCCGTTA	2129
Db	2295	aactaagctggaacttatacttctaagaacatctataccctcaacaaccccaactatct	2354
OY	2130	TGCTTCGGTAAACCTTAATGTAAACGTCGGGATCTCTCAGAAGGTTTCTTCAAAATCC	2189
Db	2355	gagattagaagagcaaaagttctcgtatcagctatagaaaaatccctlaggaanaatcc	2414
OY	2190	TCTTATGATTTTTCATTTTGTGTGCTTAATGTCATGTCACCAACGATATGAANAACGA	2249
Db	2415	ccta---gaccttgaatgccaagttctgtcagccatcagaacaacgataatgaaacga	2471
OY	2250	CTACGCAAAATTTCCCTAATGGTGAANAACAGCTGGANAACAATTTGTGGCGATATGAGATG	2309
Db	2472	ctataactcattcgcagaacatccgaaggtctcgtgaaacagcgtgtatagctcgtgatat	2531
OY	2310	CGAGAGGAGCATGCCCTCTATTTGGTATTTTGAGAAAGGAGACATTTTCCAGAGTCCATCC	2369
Db	2532	cggcctagacctctcttctgtctcttccaaccacatacctcttccaagaacctatacc	2591
OY	2370	ATTTATGAACCTACAAATTAATTAATGCTTAATCAAGGAGATTTCAAGAGACATCGACA	2429
Db	2592	acagatgaagtcgaatagttatgatacaaaaatagcttcttcgaaagctcattgta	2651
OY	2430	TGGCGATGATTTGATATGGAGATTAAACATGATTTTGTACCTTAAGCATACGCTT	2489
Db	2652	tggcgttggtttttagtatctggaagcgtcttaacccctcgtatctcgtggtggtggaatt	2711
OY	2490	TGAGAGCTGGCACTTTCCTAGAGATGTACTTAATGACCTTAACTTCTCCTATATCTGTA	2549
Db	2712	cgtagcagggagatatcggagatactccaacattagctctcagatattcttgttccga	2771
OY	2550	TATTTTCCGTAAGATCCCTCATGTGAGCTGCTGCTGTGATTTAGCGAGACTCCTGCT	2609
Db	2772	tgtctatcgttaacaatcccacaatctacagcgcactctgtgagtagcccaagactcttgaa	2831
OY	2610	TGTTCCGGAGCAACACTATCAAGACATCTTTTGAAGGAGATGGAAGGGTGGGTATCA	2669
Db	2832	aattcgcggttggaactcttccaagaacagcatcttactcgaagggttagcaacaactaagt	2891
OY	2670	CTTTAAGCACTATCTACAGCTCTTATGTGAGAGAGATATAGATCCGCCCCCATGTCTAG	2729



QY	1640	CTGCTGAGAGATTCCATACCGGTAAAGCCCATTTAG-----GTTTTCAAAGCAATTG	1691
Db	1677	gagctctcttccaccaccgagcctctccaagaacccgaaacccactagggatcagaagaaacg	1736
QY	1692	GAATTTAGCTTGGAGCA---GGAACTGGAAACAAAGTTGGAGAAATTTCTGTGGATTAAT	1748
Db	1737	gcagctgtctctgggcaaatgcaacatccctccaaanaataggaaagcatcaactggaacccgtac	1796
QY	1749	AAATTATTAAGCCATAGACCTGTGAAAAAGAGAAAATTTAGTTCTTAATATCTGTGGGGAA	1808
Db	1797	aggaatacttccctagctctcgagagaaagaataatccctccctaaatagcttatgaggaa	1856
QY	1809	TGCTGTAAATGTGAGATTCCTTAATGAGAGTTCCAAAGACCCATGCATGAGACTTAACAGAC	1868
Db	1857	cttatatagataatagctctgatacaataagcttatagaaacaagtcagtgaggagccttt	1916
QY	1869	AGATCGAGGCGCTGTGATGATGATGAATTTGGAAATTTCTTCATGTATGTGGCTCCGAAGA	1928
Db	1917	tggagctgagactatggtccttcaaggaaattcgaaattctctctataagattctatgcccac	1976
QY	1929	CAATATTAAGGTACCGTCATTAACACGCGGTGATATGTTCATCTGAATTAATGAGATCAC	1988
Db	1977	ccgcagatggttccgcacataacagcggggttatgcactaggatcacagaaacaatccc	2036
QY	1989	ACCTAACCATATATCTTCATGCGATTTTCCCACTCTTTACTATAGACAAAGACTATAGC	2048
Db	2037	tgcgcaggaatcaagcttactcttgcctctgcagactcttcttagagatactgcaatcatat	2096
QY	2109	CCTAGGGAATATTTTCCGTTATGCTTCGCGTAACCCCTATGTAAACGTGCGGATTTCTTC	2168
Db	2157	gctctcgacatccgaatctccctctctgaggaaagcaacccgagctccctggtgctctc	2216
QY	2169	AAGAAGTTTCTTCAAAAATCCTCTATGATTTTCAATTTTGTGCTTATGTGATATGC	2228
Db	2217	tgaatctcccaatcatctcttctatcgc---ttcgatgctaattcaattcatctctac	2273
QY	2229	CACCAATGATATGAATAAACGACTAGCCAAATTTCCCTATGGGAAAAACAGCTGGAGAA	2288
Db	2274	agaacaacacaagaagacatatataaccgataacccatacaacaagggtctctggagaa	2333
QY	2289	CAATTTGTTGGGCTATAGAGTGGGAGGAGCAATGCCCTCATTTGTGATTGTGAACGGAG	2348
Db	2334	cgaagctctctgagagatcttggagctagcctgct---tttgtattctcgttcogta	2390
QY	2349	ACTTTTCAAAGGTGCCATCCATTATGAACTACAAATTAATTTATGCTTATCAAGGAGA	2408
Db	2391	actctbgaagaagtcgaacctcttgcataagtaacgataatactatctatgcatcagaaga	2450
QY	2409	TTTTCAAAGACACACAGATGAGAGCGCGGTAGATTAGTAAATGGAGATTACATGATTTTC	2468
Db	2451	cttctaaagacgtaacgctgaagagacgcgtcttcaataaagcgagcttacaacgtaga	2510
QY	2469	TGTACTCTTAGGCATAGCGTTGTAGAGAGCTGGCACTTTCTGAGATGTACTCTATGACTT	2528
Db	2511	gattccataagcgctacccttgcgaagagacccaataatcagaaaaaggaacttaagctc	2570
QY	2529	TAGTTTCTCTATATTTCTGATATTTTCCGTNAGGATCCCTCATGTGAGACTGCTGTGT	2588
Db	2571	tactctatglatatactcgatgcttaccagcgcaatccctaattgttcaaaacttccctaat	2630
QY	2589	GATTAGGGAGACTCCTGGCTGTTCCGGCACACACAGTATACAGACATGCTTTTGTAGG	2648
Db	2631	agccaagagatcactaactgtagcgctatagtaaccaactcgacagcaaaagtttctcgt	2690
QY	2649	GAGTGAAGCGGTGGGTATACCTTTTAAACGACTATACTAGCTCTATTATTCAGAGAAATAT	2708
Db	2691	tcgtgtcgtgaaacattcttccaagtgaaccccccaacttgaaatcttgcgtcaatctcgttt	2750
Y	2709	AGAAATGCGCCCATGCTAGGAATTAATATATTAATATGCTGTGAACAAATTTGCTTT	2765

DB	2751	tgaagtagcagaattcttcacgaaatacataacaacaacctagctctaagtgtttt	2807
<hr/>			
RESULT	12		
AAA30848	ID	AAA30848	standard; DNA; 2784 BP.
XX	AA	AAA30848;	
AC	AA	AAA30848;	
XX	DT	29-AUG-2000	(first entry)
DE	XX	Chlamydia antigen CPN100634	gene open reading frame.
XX	KW	Chlamydia antigen; diagnosis; infection; community acquired pneumonia;	
KM	RV	therapy; upper respiratory tract disease; bronchitis; sinusitis;	
KW	ds.	asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;	
OS	XX	Chlamydia pneumoniae.	
FH	Key	Location/Qualifiers	
FT	CDS	1..2784	
FT		/tag=	
FT		/transl_except= (pos:478..479; aa:Ser)	
FT		/transl_except= (pos:519..520; aa:Ile)	
FT		/transl_except= (pos:2069..2073; aa:Leu)	
FT		/partial	
FT		/product= Chlamydia antigen CPN100634	
FT		/note= "no stop codon given"	
PN	WO	2000032794-A2.	
PD	08-JUN-	2000.	
XX	01-DEC-	1999;	99MO-CA01147.
PE	01-DEC-	1998;	98US-0110339.
PR	01-DEC-	1998;	98US-0110340.
PR	01-DEC-	1998;	98US-0110427.
PR	01-DEC-	1998;	98US-0110428.
PR	01-DEC-	1998;	98US-0110438.
PA	(CONN - )	CONNAUGHT LAB LTD.	
PI	Murdin AD,	Oomen RP,	Wang J;
DR	WPI:	2000-412339/35.	
P-PSDB:	MAV90236.		
XX	Nucleic acids encoding polypeptide antigens from Chlamydia useful for		
PT	preventing, diagnosing and treating diseases such as community acquired		
PT	pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset		
PT	asthma -		
XX	Claim 2:	Page 140-141; 174pp;	English.
PS	This sequence encodes a Chlamydia antigen of the invention, designated		
XX	CPN100634. The nucleic acids (and their complementary sequences) may be		
CC	used as diagnostic agents for detecting the presence of nucleic acids		
CC	encoding Chlamydia antigens in samples according to standard methods,		
CC	and therefore, for diagnosing Chlamydia infections. For example, they may		
CC	be used as primers and probes for diagnostic polymerase chain reaction		
CC	(PCR) assays. Antisense sequences may be used to down regulate		
CC	expression of the proteins and may be used to treat infections. The		
CC	nucleic acids may also be used to produce the protein antigens they		
CC	encode according to standard recombinant DNA methodologies. The		
CC	proteins may then be used as antigens for the production of antibodies		
CC	(i.e. as vaccines) for preventing infection by Chlamydia. The		
CC	antibodies may also be used as diagnostic reagents for detecting		
CC	infections. Chlamydia is a pathogen implicated in the development of		
CC	(for example) community acquired pneumonia, upper respiratory tract		
CC	disease (especially bronchitis and sinusitis, asthmatic bronchitis,		







```

Db 100225 gttctgaaacttaattagcttgccttcccaactcttctgagcgataaagattct 100284
QY 2048 CGTTTCCACACAGATACAGATATGATTTAGATGCTATCTATCAATATCAACT 2107
Db 100285 tagtcgctaaatacactactgatactatcgagagctctctctatcca 100333
QY 2108 CCCATGGAATATTTCCGTATGCTTCGCGTACCCCTAATGTAAGTCGGATTTCT 2167
Db 100334 acacattacgaatgtagtgggttcataagtgctctcttgataaacttctgctt 100393
QY 2168 CAGAAGGTTCTTCCAAATCTCTTATGATTTTTCATTTTGTGCTTATGCTATG 2227
Db 100394 ga-----gcatataacccctcgctttagaagggcagctcgctatagccacg 100440
QY 2228 CCACCAATGATATGATAACAGACTACGCAAAATTCCTATGTTGAAAAAGCGTGGAA 2287
Db 100441 tcagtaattgctcgaagaaagatactcgtatcccgagtgaaaggtctctg999ga 100500
QY 2288 ACAATTTGTTGGCTATGAGTGCAGAGGAGCATGCTCTATTTGTATTTGAGAACGAA 2347
Db 100501 ataagcttttaacatgatgttgagag---ctctctcatctctatccgaatacctgc 100557
QY 2348 GACTTTCCAAAGTGCCATCCCATTTATGAACATCAATTAATTATGCTTATCAGGAG 2407
Db 100558 atgttlttgatacctatgctctccatacaaaactgaaactgacatacgtacagaca 100617
QY 2408 ATTTCAAAGAGACACTGCAGATGCGCTAGATTAGTAATGAGGATTTAACATCGATTT 2467
Db 100618 gcttctcgagaaaggtacagaagaagatcttltgataagaacactctcaattat 100677
QY 2468 CTGATCTCTAGCATACGCTTTTGAGAGCTGCGACCTTCTCAGAGATGATCTATGACT 2527
Db 100678 ctctgctatagaggtgagatlttgagaagttcctcgttctgataagcttcttattg 100737
QY 2528 TTAGTTCTCTATATTCCTGATATTTTCCGTAAAGATCCCTCATGTGAAGCTCTGTG 2587
Db 100738 tgcattatctcattgctcctgattctatccgcaatgataccaaatgacactagcact 100797
QY 2588 TGATTTCGCGAGACACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2647
Db 100798 taatcagcgagagctctcttggaactatcctaactaactaagcagagccttgcaag 100857
QY 2648 GGAAGTGAACGGGTGCTGCTATCTTAAACATATGAGCTGCTGCTGCTGCTGCTGCT 2707
Db 100858 tgcgtgcagcagcactacgactcctcctcctatggttgagtgctgcgcaggttgc 100917
QY 2708 TAGAATGCCGCCGCCATGCTAGGAATATATATTAACCTGTGGAAGCAATTTGTTT 2767
Db 100918 ttgagagtcgtgagtcctcagcagattataatgtagatcttg999gtaagttcca 100977
QY 2768 AGAAG 2772
Db 100978 agag 100982

```

```

RESULT 14
AA06817
ID AAX06817 standard; DNA; 2815 BP.
XX
AC AAX06817;
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp5 DNA.
XX
KW Omp5; outer membrane protein 5; surface exposed protein; antigen;
XX infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
OS Chlamydia pneumoniae.
XX
FH Key Location/Qualifiers
CDS 1..2787

```

```

FT /*tag- a
XX
PN MO9858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98MO-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
XX (CHRI/) CHRISTIANSEN G.
XX
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mygind P;
XX
DR WPI: 1999-105610/09.
XX
DR P-PSDB: AAW88418.
XX
PS
XX
CC This DNA sequence codes for the novel 97.2 kDa surface exposed
CC protein Omp5 (see AAW88418) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see AAX06816-27) encoding
CC Omp4-Omp5 proteins (see AAW88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
CC the other, and encode polypeptides of about 89.6-100.3 kDa and
CC about 56.1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp5 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.
XX
SQ Sequence 2815 BP; 790 A; 575 C; 626 G; 824 T; 0 other;

Query Match 7.6%; Score 231.6; DB 20; Length 2815;
Best Local Similarity 48.7%; Pred. No. 7,7e-56;
Matches 997; Conservative 0; Mismatches 959; Indels 90; Gaps 10;

```

```

QY 787 ACTTCAGGAAGTAGTACTCCAGTCTCTATTGTGACTTCTCTGACATAAAGCTTAGTC 846
Db 775 atgagagagctcttcttgagatgctcgatgtaccatactcgtgagtagtaact 834
QY 847 TTGGAAGAAGAACCTATTCATATGCGTGGGAGCCATTATCTCTAGGAACCTTAGC 906
Db 835 ttctcaggaaccacagctgtagtaagtgagcagcattatgtaagaagcttaacgt 894
QY 907 TCTTCAGGAGGTC-----TACTCTATTATTCATATATATATATATGCAATTTGCA 960
Db 895 gcttcg999g999g999ggtatcctcttctacacataagtcacaggtaccactgca 954
QY 961 AATTTCAGTGGAGCTATTTGCAATGATGAGAGGAGATCAGTTTATCAGCAGAGAAA 1020
Db 955 ggtataagtgtagcattcttatactgcaagcgtgagagtgtagtcttccagagaaga 1014
QY 1021 GGAACATTATTCATTCAGAGAACGAGACGACTTACCGTTTGTG----- 1065

```







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 03:26:18 ; Search time 268.88 Seconds  
(without alignments)  
2148.831 Million cell updates/sec

Title: US-09-446-677B-5  
Perfect score: 3052  
Sequence: 1 ATCGCATTTTCGCTGCGG.....GTTCCTTGAGCTCTTGCGCA 3052

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents: NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCITUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/Backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39.8	1.3	246240	2	US-08-724-394A-20
2	39.8	1.3	246240	2	US-08-724-394A-21
3	39.8	1.3	246240	2	US-08-724-394A-22
4	36.4	1.2	12124	1	US-08-181-271A-36
5	36.4	1.2	12124	1	US-08-449-315-36
6	36.4	1.2	12124	1	US-08-444-803-36
7	36.4	1.2	12124	1	US-08-449-043-36
8	36.4	1.2	12124	1	US-08-456-265A-36
9	36.4	1.2	12124	1	US-08-455-416-36
10	36.4	1.2	12124	1	US-08-455-244-36
11	36.4	1.2	12124	1	US-08-454-876-36
12	36.4	1.2	12124	2	US-08-457-364-36
13	36.4	1.2	12124	2	US-08-456-262-36
14	36.4	1.2	12124	2	US-08-456-240-36
15	36.4	1.2	12124	2	US-08-455-736-36
16	36.4	1.2	12124	2	US-08-971-217-36
17	36.2	1.2	4314	2	US-08-199-780-2
18	36.2	1.2	4314	2	US-08-318-650-2
19	34.8	1.1	2233	1	US-08-496-631-1
20	34.4	1.1	3666	2	US-08-682-517-13
21	34.4	1.1	3666	2	US-08-682-517-14
22	34.4	1.1	4197	2	US-08-682-517-7
23	34.4	1.1	2259	6	US-08-682-517-8
24	33.8	1.1	2259	6	5185254-3
25	33.4	1.1	7218	1	US-08-232-463-14
26	33	1.1	1683	3	US-09-192-545-1
27	32.6	1.1	3580	4	US-09-081-345-1

28	32.6	1.1	6973	1	US-08-478-370-1	Sequence 1, Appli
29	32.6	1.1	9542	4	US-08-968-685A-9	Sequence 9, Appli
30	32.4	1.1	2273	3	US-08-714-918-40	Sequence 40, Appli
31	32.4	1.1	2273	4	US-09-265-315-40	Sequence 40, Appli
32	32.4	1.1	2273	4	US-09-265-315-40	Sequence 40, Appli
33	32.4	1.1	2273	4	US-09-266-417-40	Sequence 40, Appli
34	32	1.0	1962	4	US-08-714-918-18	Sequence 18, Appli
35	32	1.0	1962	4	US-09-265-315-18	Sequence 18, Appli
36	32	1.0	1962	4	US-09-265-315-18	Sequence 18, Appli
37	32	1.0	1962	4	US-09-266-417-18	Sequence 18, Appli
38	32	1.0	7218	1	US-08-232-463-14	Sequence 14, Appli
39	31.8	1.0	3098	2	US-08-818-070-1	Sequence 1, Appli
40	31.8	1.0	3098	2	US-08-723-585-1	Sequence 1, Appli
41	31.8	1.0	11832	2	US-08-416-603-1	Sequence 1, Appli
42	31.6	1.0	1469	3	US-09-074-912-1	Sequence 1, Appli
43	31.6	1.0	2031	6	5495001-8	Sequence 1, Appli
44	31.4	1.0	1079	1	US-08-454-196-16	Sequence 16, Appli
45	31.4	1.0	1079	3	US-09-064-033-16	Sequence 16, Appli

#### ALIGNMENTS

RESULT 1  
US-08-724-394A-20/c  
Sequence 20, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Krommal, Gregory S.  
APPLICANT: Laufer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchihashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
TITLE OF INVENTION: Sequences and Antibodies Thereeto  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESS: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.COMTIG"  
US-08-724-394A-20



	Query Match	Similarity	1.3%	Score 39.8	DB 2	Length 246240			
	Best Local	Similarity	50.3%	Pred. 1.2					
	Matches	98	Conservative	0	Mismatches	97	Indels	0	Gaps
OY	1131	AGATTTTATGATCTTACTTCTGAAACGAGGCGTCCACCATTTGAATATCAACGG	1190						
Db	54152	AGAGTTTATGATACATTTTGTACTGTATGTGATGTTTCTTTCATTTAAATAAT	54093						
OY	1191	AGATCTTAAATATAGAGTACACAGGAGCACCATCTTTTCTGGAGAAAGAGTCTAGC	1250						
Db	54092	CAACCGTCAATTTAAATGCTAGATTGTTCTGCTCTTTTGTATTACACAGCATTAATCT	54033						
OY	1251	AAACGATCTAGAGATTTTAAATCTACATCCCTCAGAAACGTCAACCTGTCTGCAGATA	1310						
Db	54032	AAAATATTTGTAGTATTTTATATCTCCCTGCCTAAAGCAATTAACATCTATCATCAGCAGA	53973						
OY	1311	CTTAGTTATTTAAAG	1325						
Db	53972	CTAGACCAATTAATAA	53958						

```

RESULT      2
US-08-724-394A-21/c
Sequence 21, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Laufer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchinashi, Kenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESS: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Filts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H. CONTIG"
US-08-724-394A-21

```

Query Match	1.3%;	Score 39.8;	DB 2;	Length 246240;
Best Local Similarity	50.3%;	Pred. No. 1.2;		
Matches	98;	Conservative	0;	Mismatches 97; Indels 0; Gaps 0;
QY 1131	AGAAATTTATGATCCTATTACTCTCGAAGCAGATGGGCTCAACCAATGGAATATCAACGG	1190		
Db 54152	AGAGTTTATATATACATTTTGTATGCTATGATGTGATGTTTCTATTTATTTAAATAAAT	54093		
QY 1191	AGATCTCTAAATAATAGAGTACACAGGAGCACAATCTCTTTCTGGAGAAAGAGTCTAGC	1250		
Db 54092	CACCGTGTCAATTTAATGATGATGTCTGTGCTCTTTTGGATTGACACAGTCATTAATCT	54033		
QY 1251	AAACGATCTAGGAGTTTAAATCTCAATATCCCTCAGAACGTCAACCTGTCTGCAGGATA	1310		
Db 54032	AAAATATATGTGTATTTTATATCTCCCTGCCTAAAGCATAAACAATCTAATCAGCAGA	53973		
QY 1311	CTTAGTTTATTAAGA	1325		
Db 53972	CTAGACACATATAAAA	53958		

```

RESULT 3
US-08-724-394A-22/c
? Sequence 22, Application US/08724394A
? Patent No. 5872237
?
GENERAL INFORMATION:
? APPLICANT: Feder, John N.
? APPLICANT: Kronmal, Gregory S.
? APPLICANT: Laufer, Peter M.
? APPLICANT: Ruddy, David A.
? APPLICANT: Thomas, Winston
? APPLICANT: Tsuchinashi, Zenta
? APPLICANT: Wolff, Roger K.
? TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
? TITLE OF INVENTION: Sequences and Antibodies Thereto
? NUMBER OF SEQUENCES: 31
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
? STREET: Two Embarcadero Center, 8th Floor
? CITY: San Francisco
? STATE: CA
? COUNTRY: USA
? ZIP: 94111-3834
?
COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/724,394A
? FILING DATE: 01-OCT-1996
? CLASSIFICATION: 536
? ATTORNEY/AGENT INFORMATION:
? NAME: Fitts, Renee A.
? REGISTRATION NUMBER: 35,136
? REFERENCE/DOCKET NUMBER: 017957-000100
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-576-0200
? TELEFAX: 415-576-0300
? INFORMATION FOR SEQ ID NO: 22:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 246240 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: not relevant
? TOPOLOGY: not relevant
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: misc feature
? LOCATION: 1..246240
? OTHER INFORMATION: /note= "HLA-H.COMTIG"
?
US-08-724-394A-22

```



Db 10108 AA 10107

```
RESULT 5
US-08-449-315-36/c
Sequence 36, Application US/08449315
Patent No. 5650505
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,315
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ. ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 12124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Cucumis sativus
INDIVIDUAL ISOLATE: Cucumber Chitinase Genomic DNA
IMMEDIATE SOURCE:
CLONE: pBSclucchrht5
US-08-449-315-36
```

Query Match 1.2%; Score 36.4; DB 1; Length 12124;  
Best Local Similarity 50.0%; Pred. No. 2.4;  
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

```
QY 1992 TAAGCACTATCTTCGATGCGATTTTCCCACTCTTTAGTAGACAGACGACTATGCGGT 2051
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10288 TAAATTATTACAAATCACTTGAGATTATTCATAGCTTAATTAACCTTAGAGGCT 10229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2052 TTCACACAGCAATGATGAGATGATTTAGATGCTATCTATCAATATCAACCTCCCT 2111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10228 TCCCTTCAGCTGATATATGCTGCTTTAATGCGATCGCTTAGCATTGTCAACGCTT 10169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2112 AGGGAATATTTCCGTTATCTTCGCGTAAACCTAATGTAAAGCTCGGATTCCTCAAG 2171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10168 ACTGCATATACATCACCCTCCATAGTTGGAATTAAGCTTTAATGTTGGAGAACCTCAGA 10109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2172 AA 2173
    ||
Db 10108 AA 10107
```

```
RESULT 6
US-08-444-803-36/c
Sequence 36, Application US/08444803
Patent No. 5654414
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
```





```

1      FILING DATE: 21-DEC-1990
2      PRIOR APPLICATION DATA:
3      APPLICATION NUMBER: US 07/425,504
4      FILING DATE: 20-OCT-1989
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: US 07/848,506
7      FILING DATE: 6-MAR-1992
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: US 07/768,122
10     FILING DATE: 27-SEP-1991
11     PRIOR APPLICATION DATA:
12     APPLICATION NUMBER: US 07/580,431
13     FILING DATE: 7-SEP-1990
14     PRIOR APPLICATION DATA:
15     APPLICATION NUMBER: US 07/368,672
16     FILING DATE: 20-JUN-1989
17     PRIOR APPLICATION DATA:
18     APPLICATION NUMBER: US 07/329,018
19     FILING DATE: 24-MAR-1989
20     PRIOR APPLICATION DATA:
21     APPLICATION NUMBER: US 08/045,957
22     FILING DATE: 12-APR-1993
23     ATTORNEY/AGENT INFORMATION:
24     NAME: Meigs, J. Timothy
25     REGISTRATION NUMBER: 38,241
26     REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727/DIV10
27     TELECOMMUNICATION INFORMATION:
28     TELEPHONE: (919)541-8587
29     TELEFAX: (919)541-8669
30     INFORMATION FOR SEQ ID NO: 36:
31     SEQUENCE CHARACTERISTICS:
32     LENGTH: 12124 base pairs
33     TYPE: nucleic acid
34     STRANDEDNESS: single
35     TOPOLOGY: linear
36     MOLECULE TYPE: DNA (genomic)
37     ORIGINAL SOURCE:
38     ORGANISM: Cucumis sativus
39     INDIVIDUAL ISOLATE: Cucumbr Chitinase Genomic DNA
40     IMMEDIATE SOURCE:
41     CLONE: pBScucnchrts
42     US-08-456-265A-36

```

FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Cucumis sativus  
INDIVIDUAL ISOLATE: Cucumber Chitinase Genomic DNA  
IMMEDIATE SOURCE:  
CLONE: pBSucchrchts  
US-08-455-416-36

Query Match 1.2%; Score 36.4; DB 1; Length 12124;  
Best Local Similarity 50.0%; Pred. No. 2.4;  
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1992 TAAGCACTATATCTCGATGCGATTTCCCACTTTAGTAGAGACAGACTATGCGGT 2051  
DB 10288 TAATTAATTAACTTGGAGTTTATTCATAGCTTAATTAATAAAGTAACTTAGAGCT 10229  
QY 2052 TTCCACACAGCATTAAGATGATTTAGATGCTATCTATCAATATCAACCTCCCT 2111  
DB 10228 TCCCTTCAGCTGATATATGCTTCTTAATGATGATGCTATGATGCTATGATGCT 10169  
QY 2112 AGGCAATATTTCCGTTATCTGCGTACCTTAATGTAAGTCGCGATTCCTCAAG 2171  
DB 10168 ACCTCAATACATACCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 10109  
QY 2172 AA 2173  
DB 10108 AA 10107

RESULT 10  
US-08-455-244-36/C  
Sequence 36, Application US/08455244  
Patent No. 5789214  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedlich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Urdes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Sherilca C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)



ORIGINAL SOURCE:  
ORGANISM: Cucumis sativus  
INDIVIDUAL ISOLATE: Cucumber Chitinase Genomic DNA  
CLONE: pbscucchrht5  
US-08-455-244-36

Query Match 1.2%; Score 36.4; DB 1; Length 12124;  
Best Local Similarity 50.0%; Pred. No. 2.4;  
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1992 TAAGCACTATACCTGATGCAATTTCCCACTCTTAGTAGAGACAGACATGCGGT 2051  
DB 10288 TAATTTATTAACATCTTGGAGTTTATTCATGCTTAATTAACATACTTAGAGCT 10229  
QY 2052 TTCACACAGCAATACAGAAATGATTAGATGCTATCTATCAATATACACCTCCCT 2111  
DB 10228 TCCCTCAGCTGATATATGCTGCTTTAATGCGATCGCTAGACATGTCAACGCTT 10169  
QY 2112 AGGGAATATTTCCGTTATCTCGCCTAACCTAATGTAAAGTCGGATTCTCTCAG 2171  
DB 10168 ACTCCATACATCATCTCCCTCATGTTGAATAGCTTAATGTTGGAAGAACTTCAGA 10109  
QY 2172 AA 2173  
DB 10108 AA 10107

RESULT 11  
US-08-454-876-36/C  
Sequence 36, Application US/08454876  
Patent No. 5804693

GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedlich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Umes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericea C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY CORPORATION  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/454,876  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94

APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Cucumis sativus  
INDIVIDUAL ISOLATE: Cucumber Chitinase Genomic DNA  
IMMEDIATE SOURCE:  
CLONE: pbscucchrht5  
US-08-454-876-36

Query Match 1.2%; Score 36.4; DB 1; Length 12124;  
Best Local Similarity 50.0%; Pred. No. 2.4;  
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1992 TAAGCACTATACCTGATGCAATTTCCCACTCTTAGTAGAGACAGACATGCGGT 2051  
DB 10288 TAATTTATTAACATCTTGGAGTTTATTCATGCTTAATTAACATACTTAGAGCT 10229  
QY 2052 TTCACACAGCAATACAGAAATGATTAGATGCTATCTATCAATATACACCTCCCT 2111

DB 10228 TCCCTTACGCTGATATATGCTGCTTTAATGAGCATGCTGATGCAACGCGCTT 10169  
QY 2112 AGGGAATATTTTCGGTATGCTGCGGTACCCCTAATGTAAGTGGGATTCCTCAAG 2171  
DB 10168 ACTCCATACATCCTCCTCCATAGTTGGATATAGCTTTATGTTGGAGAACTTCAGA 10109  
QY 2172 AA 2173  
DB 10108 AA 10107  
RESULT 12  
US-08-457-364-36/C  
Sequence 36, Application US/08457364  
Patent No. 5847258  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Meyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Sherilca C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,364  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 800  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Cucumis sativus  
INDIVIDUAL ISOLATE: Cucurbit Chitinase Genomic DNA  
IMMEDIATE SOURCE:  
CLONE: pBSucucnrchts  
US-08-457-364-36  
Query Match 1.2%; Score 36.4; DB 2; Length 12124;  
Best Local Similarity 50.0%; Pred. No. 2.4;  
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;  
QY 1992 TAAGCACTATATCTTCGATGAGCATTTTCCCACTCTTTAGTAGAGACAGGACTATGCGGT 2051  
DB 10288 TAATTTATTACAAATACCTTTGGAGTTTATTCATGACTTTAATTAACCTAAGAGAGCT 10229  
QY 2052 TTCCACACAGCAATACAGATGATTTAGATGCTATCTATCACTATATCAACCTCCCT 2111  
DB 10228 TCCCTTACGCTGATATATGCTGCTTTAATGATGATGCTGCTGATGCAACGCGCTT 10169  
QY 2112 AGGGAATATTTTCGGTATGCTGCGGTACCCCTAATGTAAGTGGGATTCCTCAAG 2171  
DB 10168 ACTCCATACATCCTCCTCCATAGTTGGATATAGCTTTATGTTGGAGAACTTCAGA 10109  
QY 2172 AA 2173  
DB 10108 AA 10107  
RESULT 13  
US-08-456-262-36/C  
Sequence 36, Application US/08456262  
Patent No. 5851766  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.

APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
TITLE OF INVENTION: SHERICCA C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,262  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/PI/GCC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: Cucumis sativus  
INDIVIDUAL ISOLATE: Cucumber Chitinase Genomic DNA  
IMMEDIATE SOURCE:  
CLONE: pBscucchrht5  
US-08-456-262-36  
Query Match 1.2%; Score 36.4; DB 2; Length 12124;  
Best Local Similarity 50.0%; Pred. No. 2.4;  
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;  
QY 1992 TAAGCACTATCTGATGCGATTTTCCCACTTTAGTAGAGACAAGACTATGGGT 2051  
DB 10288 TAATATTACAAATCACTTGAGTTATTCATAGCTTAATTAACCTTAGAGACT 10229  
QY 2052 TTCCAAACAAGATATAGATATGATATGATCTATCATATATCAACCTCCCT 2111  
DB 10228 TCCCTTCAGCTGATATATGCTGCTTTATATGCGATCGCTGACCATTTGCAACGCTT 10169  
QY 2112 AGGGAATATTTCCGTTATGCTTCCGTAACCCCTAATGTAACGCTGGGATTTCTCAAG 2171  
DB 10168 ACTCCATACATCATCTCCCTCATGTTGGAATAGCTTTAATGGTTGGAAGACTCAGA 10109  
QY 2172 AA 2173  
DB 10108 AA 10107  
RESULT 14  
US-08-456-240-36/c  
Sequence 36, Application US/08456240  
Patent No. 5856154  
GENERAL INFORMATION:  
APPLICANT: Rivals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
TITLE OF INVENTION: SHERICCA C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,240  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/Pl/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12124 base pairs  
type: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

? MOLECULE TYPE: DNA (genomic)  
 ? ORIGINAL SOURCE:  
 ? ORGANISM: Cucumis sativus  
 ? INDIVIDUAL ISOLATE: cucumber Chitinase Genomic DNA  
 ? IMMEDIATE SOURCE:  
 ? CLONE: pBscuchricht5  
 ? US-08-456-240-36

Query Match	1.28;	Score 36.4;	DB 2;	Length 12124;
Best Local Similarity	50.08;	Pred. No. 2.4;		
Matches 91;	Conservative 0;	Mismatches 91;	Indels 0;	Gaps 0;

Oy    1992 TAAGACATACTTCGATGGCATTTTCCCACTTTAGTAGACAAGACTAAGCGGT 2051  
||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 10288 TAAATTAACAATCTTGGAAGTTATCATAGCTTTAAATTAAAACATAACTTAGGAGCT 10229

Qy 2052 TTCACCAAGAAATACAGATGATTTTAGGATCGTATCTCTATCATATNACAACCTCCCT 2111  
| | | | | | | | | | | | | | |  
Db 10228 TCCCTTCAGCTGATATATGCTGTCTTTAATGGCATCCGCTGTAGCCATTGCAAACGCCTT 10169

Oy	2112	AGGGAAATTTTCCGTTATGCTTCGCCGTAAACCCTAATGTAAACGTCGGGATTCTCTCAAG	2171
Db	10168	ACTCCATAACATCAGCTCCCTCCATGTTGGAATAAGCTTAATGGTTGGAAGAAGACTTCAGA	10109

Oy	2172	AA	2173
Db	10108	AA	10107

RESULT 15  
US-08-455-736-36/c

Patent No. 5880328  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
ATTORNEY: Alexander, Daniel C.

APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
ADDRESS: Cedar, Robert W.

APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Montoya, Mary D

```

;
APPLICANT: Neuhaus, Jean-Marc
;
APPLICANT: Payne, George B.
;
APPLICANT: Spertison, Christoph
;
APPLICANT: Toffen-D
;

```

APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericca C.  
APPLICANT: Williams, Shericca C.

; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
 ;  
 ; NUMBER OF SEQUENCES: 106  
 ;  
 ; CORRESPONDENCE ADDRESS:  
 ;  
 ; ADDRESSEE: CITRA-GENETIC CORPORATION

STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA

```

; ZIP: 10532
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPARED: IBM PC compatible

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: ITS/08/455 736
;

```

```

: FILING DATE: 31-MAY-1995
:
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: 08/181 271
:

```

FILING DATE: 13-JAN-1994  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ. ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: Cucumis sativus  
INDIVIDUAL ISOLATE: Cucurbit Chitinase Genomic DNA  
IMMEDIATE SOURCE:  
CLONE: pBSucchrht5  
US-08-455-736-36

Query Match 1.2%; Score 36.4; DB 2; Length 12124;

Best Local Similarity 50.0%; Pred. No. 2.4;

Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1992 TAAGCACTATCTGCGATTTTCCACACTTTAGTAGAGACAAGCATATGCGGT 2051  
DB 10288 TAATTTATTAATACCTTTGAGATTATTCATACCTTTAATTAACCTTAGAGGCT 10229  
QY 2052 TTCACACACGATACGATGATATTAGATGCTATCTATCAATATACACCTCCCT 2111

DB 10228 TCCTTCAGCTGATATATGCTGTCTTTAATGCGATCGCTAGCCATGTCAACGCTT 10169  
QY 2112 AGGGAATATTTTCCGTTATGCTTGGGTAACCTATGTAAAGCTGGGATTTCTCAAG 2171  
DB 10168 ACTCCATTAACATCCTCCTCATAGTTGGAATTAAGCTTTAATGTTGGAAGAACTTCAGA 10109  
QY 2172 AA 2173  
DB 10108 AA 10107

Search completed: October 2, 2001, 03:44:51  
Job time: 51284 sec

**THIS PAGE BLANK (USPTO)**

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 03:06:19 ; Search time 10300.6 Seconds  
(without alignments)  
2800.826 Million cell updates/sec

Title: US-09-446-677B-5  
Perfect score: 3052  
Sequence: 1 ATGGGATTTCGCTCTGGCG.....GTCTCTGACGTCTTGCA 3052

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1022815 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_esthum29:\*  
64: em\_esthum30:\*  
65: em\_esthum31:\*  
66: em\_esthum32:\*  
67: em\_esthum33:\*  
68: em\_esthum34:\*  
69: em\_esthum35:\*  
70: em\_esthum36:\*  
71: em\_esthum37:\*  
72: em\_esthum38:\*  
73: em\_esthum39:\*  
74: em\_esthum40:\*  
75: em\_esthum41:\*  
76: em\_esthum42:\*  
77: em\_esthum43:\*  
78: em\_esthum44:\*  
79: em\_esthum45:\*  
80: em\_esthum46:\*  
81: em\_esthum47:\*  
82: em\_esthum48:\*  
83: em\_esthum49:\*  
84: em\_esthum50:\*  
85: em\_esthum51:\*  
86: em\_esthum52:\*  
87: em\_esthum53:\*  
88: em\_esthum54:\*  
89: em\_esthum55:\*  
90: em\_esthum56:\*  
91: em\_esthum57:\*  
92: em\_esthum58:\*  
93: em\_esthum59:\*  
94: em\_esthum60:\*  
95: em\_esthum61:\*  
96: em\_esthum62:\*  
97: em\_esthum63:\*  
98: em\_esthum64:\*  
99: em\_esthum65:\*  
100: em\_esthum66:\*  
101: em\_esthum67:\*  
102: em\_esthum68:\*  
103: em\_esthum69:\*  
104: em\_esthum70:\*  
105: em\_esthum71:\*  
106: em\_esthum72:\*  
107: em\_esthum73:\*  
108: em\_esthum74:\*  
109: em\_esthum75:\*  
110: em\_esthum76:\*  
111: em\_esthum77:\*  
112: em\_esthum78:\*  
113: em\_esthum79:\*  
114: em\_esthum80:\*  
115: em\_esthum81:\*  
116: em\_esthum82:\*



```
117: qb_est48:*
118: qb_est49:*
119: qb_est50:*
120: qb_est51:*
121: qb_est52:*
122: qb_est53:*
123: qb_est54:*
124: qb_est55:*
125: qb_est56:*
126: qb_est57:*
127: qb_est58:*
128: qb_est59:*
129: qb_est60:*
130: qb_est61:*
131: qb_est62:*
132: qb_est63:*
133: qb_est64:*
134: qb_est65:*
135: qb_est66:*
136: qb_est67:*
137: qb_est68:*
138: qb_est69:*
139: qb_est70:*
140: qb_est71:*
141: qb_est72:*
142: qb_est73:*
143: qb_est74:*
144: qb_est75:*
145: qb_est76:*
146: qb_est77:*
147: qb_est78:*
148: qb_est79:*
149: qb_est80:*
150: qb_est81:*
151: qb_est82:*
152: qb_est83:*
153: qb_est84:*
154: qb_est85:*
155: qb_est86:*
156: qb_est87:*
157: qb_est88:*
158: qb_est89:*
159: qb_est90:*
160: qb_est91:*
161: qb_est92:*
162: qb_est93:*
163: qb_est94:*
164: qb_est95:*
165: qb_est96:*
166: qb_est97:*
167: qb_est98:*
168: qb_est99:*
169: qb_est100:*
170: qb_est101:*
171: qb_est102:*
172: qb_est103:*
173: qb_est104:*
174: qb_est105:*
175: qb_est106:*
176: qb_est107:*
177: qb_est108:*
178: qb_est109:*
179: qb_est110:*
180: qb_est111:*
181: qb_est112:*
182: qb_est113:*
183: qb_est114:*
184: qb_est115:*
185: qb_est116:*
186: qb_est117:*
187: qb_est118:*
188: qb_est119:*
189: qb_est120:*
```

```
190: qb_est121:*
191: qb_est122:*
192: qb_est123:*
193: qb_est124:*
194: qb_est125:*
195: qb_est126:*
196: qb_est127:*
197: qb_est128:*
198: qb_est129:*
199: qb_est130:*
200: qb_est131:*
201: qb_est132:*
202: qb_est133:*
203: qb_est134:*
204: qb_est135:*
205: qb_est136:*
206: qb_est137:*
207: qb_est138:*
208: qb_est139:*
209: qb_est140:*
210: qb_est141:*
211: qb_est142:*
212: qb_est143:*
213: qb_est144:*
214: qb_est145:*
215: qb_est146:*
216: qb_est147:*
217: qb_est148:*
218: qb_est149:*
219: qb_est150:*
220: qb_est151:*
221: qb_est152:*
222: qb_est153:*
223: qb_est154:*
224: qb_est155:*
225: qb_est156:*
226: qb_est157:*
227: qb_est158:*
228: qb_est159:*
229: qb_est160:*
230: qb_est161:*
231: qb_est162:*
232: qb_est163:*
233: qb_est164:*
234: qb_est165:*
235: qb_est166:*
236: qb_est167:*
237: qb_est168:*
238: qb_est169:*
239: qb_est170:*
240: qb_est171:*
241: qb_est172:*
242: qb_est173:*
243: qb_est174:*
244: qb_est175:*
245: qb_est176:*
246: qb_est177:*
247: qb_est178:*
248: qb_est179:*
249: qb_est180:*
250: qb_est181:*
251: qb_est182:*
252: qb_est183:*
253: qb_est184:*
254: qb_est185:*
255: qb_est186:*
256: qb_est187:*
257: qb_est188:*
258: qb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.





```

/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/Note="Organ: sperm; Vector: BAC63.6; BAC Clones in E-Coli
DH10B"
BASE COUNT      240 a      147 c      134 g      394 t
ORIGIN

Query Match
Best local Similarity 55.1%; Score 41.4; DB 239; Length 915;
Matches 81; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1690 TGGAAATGACCTGGACAGAGACTGGAAACAAAGTGGAGAAATCTTCGTGGATAATA 1749
      || || || || || || || || || || || || || || || || || || || || ||
Db 561 TGAACACTAAATGATAGTACTGACAAATGAGAAATGAGCAAAATAGTGAACATAATA 502

QY 1750 AATTATTAAGCTTGACCTGAAAAAGAAAGAAATTTAGTCTTATATCTTGTTGGGGAAT 1809
      || || || || || || || || || || || || || || || || || || || || ||
Db 501 TATTATTAATAAAGAAATTTATTAATAAAGAGGAAATTTAGATGAGAAAGCTTACAGGAATAGT 442

QY 1810 GCTTAATGTCAGATCTTAATGACAG 1836
      || || || || || || || || || || || || || || || || || || || || ||
Db 441 AACATATTTGGTAACTGACCAATGAG 415

RESULT 5
BF290068 554 bp mRNA EST 28-NOV-2000
LOCUS EST545659 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
DEFINITION Rattus norvegicus cDNA clone KGIHN37 3' sequence, mRNA sequence.
ACCESSION BF290068
VERSION BF290068.1 GI:11221138
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 554)
AUTHORS Malek,R.L., Cho,J., Lee,Y., Karamecheva,S., Parvizi,B., Pertea,G.,
Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
Generation of ESTs from Normalized Rat Embryo, Bento Soares
Other ESTs: EST352405
JOURNAL Unpublished (2000)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.

FEATURES
source
1..554
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RGJHN37"
/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/tissue_type="mixed tissue"
/lab_host="DH5-alpha"
/Note="Vector: pT373Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROV, RBR, RKI, RLI, RPL, RLU, REM, RNU, RSP
RHE, RPC, RPN"
BASE COUNT      200 a      85 c      89 g      180 t
ORIGIN

Query Match
Best local Similarity 49.5%; Score 41.2; DB 146; Length 554;
Matches 1.3%; Pred. No. 2.7;

```

```

Matches 106; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 406 AAGAACAGGAGTGTCTATTCAAAAAATGACATATGCTTTAAACATATATAGTG 465
      || || || || || || || || || || || || || || || || || || || || ||
Db 138 AAGAACATATTAATTTTCAGATTAACAGCATTTCAATTTTCTTAGGCAATTCATTAAC 197

QY 466 CGTTTGAACAAACCAATAGACATTAAGGCGGACCTTTGTGGGGGAATGACT 525
      || || || || || || || || || || || || || || || || || || || || ||
Db 198 ACTTTCCAAATGATCAGTACAGACACAGCAAAAGCTTTAAGTATATGATTTTGT 257

QY 526 ATAGTAGCACTACGATTCGTCCTTTATCAGATGACGACCTTTTGAGAGTGT 585
      || || || || || || || || || || || || || || || || || || || || ||
Db 258 ATGTAAATATACGTGTGATACCAATCTAAGTCGTATATGATACAGGTTAGTTCAA 317

QY 586 ATCCATTTCTTCAAGTCCCTTACAGATTGCAATGA 619
      || || || || || || || || || || || || || || || || || || || || ||
Db 318 ATTGCTTCACTAGAGCAAGTATGATATATATGA 351

RESULT 6
AI408352 593 bp mRNA EST 09-FEB-1999
LOCUS EST236642 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
DEFINITION ROVES15 3' end, mRNA sequence.
ACCESSION AI408352
VERSION AI408352.1 GI:4251856
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 593)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
source
1..593
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone="ROVES15"
/clone_lib="Normalized rat ovary, Bento Soares"
/Note="Organ: ovary; Vector: pT373Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT      213 a      96 c      93 g      191 t
ORIGIN

Query Match
Best local Similarity 49.5%; Score 41.2; DB 19; Length 593;
Matches 106; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 406 AAGAACAGGAGTGTCTATTCAAAAAATGACATATGCTTTAAACATATATAGTG 465
      || || || || || || || || || || || || || || || || || || || || ||
Db 138 AAGAACATATTAATTTTCAGATTAACAGCATTTCAATTTTCTTAGGCAATTCATTAAC 197

QY 466 CGTTTGAACAAACCAATAGACATTAAGGCGGACCTTTGTGGGGGAATGACT 525
      || || || || || || || || || || || || || || || || || || || || ||
Db 198 ACTTTCCAAATGATCAGTACAGACACAGCAAAAGCTTTAAGTATATGATTTTGT 257

QY 526 ATAGTAGCACTACGATTCGTCCTTTATCAGATGACGACCTTTTGAGAGTGT 585
      || || || || || || || || || || || || || || || || || || || || ||
Db 258 ATGTAAATATACGTGTGATACCAATCTAAGTCGTATATGATACAGGTTAGTTCAA 317

```

```

QY 586 ATCCATCTTCAGTCCCTACAGATTGCAATGA 619
DB 318 ATGCTTCACTAGACGAGATTGATTATGA 351

RESULT 7
LOCUS CNS00396/c 1101 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoko Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1.1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR08K10"
/note="end : TET3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others

ORIGIN
Query Match 1.3%; Score 41; DB 219; Length 1101;
Best local Similarity 14.8%; Pred. No. 3.7;
Matches 96; Conservative 276; Mismatches 273; Indels 4; Gaps 1;

QY 496 GCGGAGCTATTAGTGGGCGGAGTGTACTAGTAGAGCACTAGCATTCGCTCTTC 555
DB 1101 GARRRGDDTDWDDTDRDKDMDTKMTWMDRRMADBDADRMWMDGATMTATW 1042
QY 556 TATCAGATGACGCCACTTTTGAGAGTGCATCCAGTCTCCGCCACAGATTGCA 615
DB 1041 WWWWATWTDWTKMWWMAATAKTDJATMTWMTAMPADAGDRGAKKDRAPADADG 992
QY 616 GTAATCAGCAGAGATTAATGATTCACAAATATCTGCCAAGATGTTCTGGAGGCT 675
DB 981 AGRROGRRKRRKKDKDDDDKKKKKAAAKAAATATWMDMDKMDKMDGAKDR 922
QY 676 TGTACTCCGATGATATGATATGATGACAGATGCTTATGTTGAGAAAT 735
DB 921 ADDDAGAKDKDDDKGKDDDDTDGTRKDDDKMDMDKATGMDATWMAATDMMW 862
QY 736 GAGGATTTGACTCTCTATAGTAAAGGAGGGGCTGTGTTGTTCTCCACTTGA 795
1 : : : : : 1 : : : : : 1 : : : : : 1 : : : : :

```

```

DB 861 GWADADWTTWDAADDDWADDRDANAMKWDAMWAGARTADRRDWDGRAGKRGARR 802
QY 796 AGTAGACTCCAGTCTCTATGACTTCTCTGACAAATAAACAGTAGTCTTGAAGA 855
DB 801 DKRRADDKRADDRDADAATWTTTTRTDIDDKMKKTDJTRMADRRWDDDDDDRD 742
QY 856 AACCATTCATTAATGGTGGCGGAGCCATTATGCTAGCAACTTACCTCTTCAGCA 915
DB 741 RAGTAGKRRWTRWKRKRKRDRDTRMDADADTDARDRRRRRGDAGDKGKRGKRRR 682
QY 916 GGTCCACTCTCTCTCTCAATATATATCATATGCAATTCGCAAAATTTAGSGTG 971
DB 681 DRATWDTDMWADAMWTTTDTDDDDDKDRRRRGARRRRRTTARAADWMTWAMDW 622
QY 972 ACCTATTCGATTAATGAGGAGGAGATCACTTATATCAGCAGAAAGCAATTAAC 1031
DB 621 AKMDKTRADRRMRMADDTTADKADBDKAKRAMRARBRBARADRRWTKGTTT 562
QY 1032 ATTCCAAAGAAACCGACGAGCTTACCGTTTGAATGCGATCCATCTTTACAAATGC 1091
DB 561 ATWTTWMAARAAMWMAWMAATTTATWTTTWTWTTTWTWTTTWTWMAWMAWMTAT 502
QY 1092 TAAATTCCTGAATTTACAGCGGAGAAATGATGCTCTATGAATTTAT 1140
DB 501 WAAMTAAAMAAAAAATAATTTTTTTTTTTTWTWMAWMAWMTATWTTT 453

RESULT 8
LOCUS CNS00715/c 1191 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR15F03 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL067600.1 GI:4945782
VERSION AL067600
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1191)
AUTHORS Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoko Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1.1191
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR15F03"
/note="end : TET3"

BASE COUNT 512 a 162 c 141 g 240 t 136 others

ORIGIN

```

Query Match	1.3%	Score 40.8	DB 219	Length 1191
Best Local Similarity	38.2%	Pred. NO. 4.3		
Matches 102	Conservative 31	Mismatches 134	Indels 0	Gaps 0
QY 2727	TAGGAATTATATAATAAAGCTGGAAGCAAAATTCGTTTATGGAAGCTTCATTGCGTCG	2786		
Db 1064	TWTAAATATTCCTWTAATCAATCTGACAKTTTCAKATCTTTATATKKKKKGTATATYTCAAAC	1005		
QY 2787	TGTGGTTCGGATCTTAACTATAAATCTGACATATGATATGAGCATATGAGTGGTTCGTCG	2846		
Db 1004	TATRTYATTTATTCGCTACTTTCATTCCKAGTAACAKTWTATATGTAAGTAATATCTAT	945		
QY 2847	AACCTTGCTGGAGATATACGACATTTTATATGCAATTAAGGAATATCTCGATATACCTGACG	2906		
Db 944	TTCTATTGAAGAAKMAAATAATACGCACTTCTTCTTATATATTTCTTAAATTAATTAATCTT	885		
QY 2907	CCCTATGACACATTCCTTATGAGGCTTCTTATTTGTCTTAACTCTGATATTTATGCAATC	2966		
Db 884	GKCAAGATKACATTTTCTTACTATATGTAAGKATTTTATTTTCTTTTATKAAAGTATC	825		
QY 2967	CTTTACGTTCTTGGTTCCTTCTGTCCTC	2993		
Db 824	TTTAAKTTTCACTTATATATATCTC	798		
RESULT 9				
AG015509/c	AG015509	654 bp	DNA	GSS
LOCUS	Homo sapiens genomic DNA, 21q region, clone: R44F3B16, genomic			08-FEB-1999
DEFINITION	Survey sequence.			
ACCESSION	AG015509			
VERSION	AG015509.1	GI:4159722		
KEYWORDS	GSS.			
SOURCE	Homo sapiens DNA, clone:R44F3B16.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 654)			
TITLE	Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.			
JOURNAL	Submitted (14-JAN-1999) to the DDBJ/EMBL/GenBank databases.			
FEATURES	Direct Submission			
REFERENCE	Masahira Hattori, Kitasato University, Department of Science, JST Sequencing Laboratory, Kitasato 1-15-1, Sagamihara 228, Japan (E-mail:hattori@hgc.jms.u-tokyo.ac.jp, Tel:0427-78-9732, Fax:0427-78-9561)			
AUTHORS	2 (bases 1 to 654)			
TITLE	Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.			
JOURNAL	Homo sapiens genomic DNA, chromosome 21q			
FEATURES	Published Only in Database (1999) In press			
source	Location/Qualifiers			
	1..654			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/chromosome="21"			
	/clone="R44F3B16"			
	/map="21q"			
BASE COUNT	143 a 136 c 109 g 265 t		1 others	
ORIGIN				
Query Match	1.3%	Score 40.4	DB 219	Length 654
Best Local Similarity	49.1%	Pred. No. 4.8		
Matches 107	Conservative 0	Mismatches 111	Indels 0	Gaps 0
QY 472	GAACAAAACCAAGTAAAGACGCGGACCTATATATGGGGCGAAGTTATCTACTACTA	531		
Db 550	GAAAGAAACAGAGGAGGAGAAACATTCAGAAATTAATGCGCAAAAATTTCTTAATCTC	491		
QY 532	GCAACATGATTCGCTCTTCTATCAGATGACGACCATTTTGGAGTGCATTCAT	591		
Db 490	TGGAAGAAATGAATCATATCATATCACTCAATGCTCTGTTATAGATTAATTTCAA	431		
QY 592	TCTTCAGGTCCTCCATACAGATTCGACGTAAATCAGCGACAGATTAAGTATGCAAAAATCT	651		

[illegible]

ACCESSION	AY1691324	GI:4966468	EST.	Zea mays.	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
VERSION	AT691324.1	GI:4966468	EST.	Zea mays.	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					
FEATURES					
source					
BASE COUNT	177 a	138 c	121 g	184 t	1 others
ORIGIN					
Query Match	1.38;	Score 40.2;	DB 23;	Length 621;	
Best Local Similarity	46.68;	Pred. No. 5.4;			
Matches 129;	Conservative	0;	Mismatches 148;	Indels 0;	Gaps 0;
QY	2066	ACACAATGATTTAGTATGATCGATCTCATCAATATATACAACTCCCTAGGAATATTTCC	2125		
Db	138	ATACAAATTTAATTAATGATACGATACGATATGATGATATCAAGATCCCATCTTTGCTTCA	197		
QY	2126	GTTATTCGTCGCGTACACCTATGTAACGTCGGGATTCCTCAAGAAGTTTCTTCAA	2185		
Db	198	ATTATTTGCTTTCTTGAACCCATCTAGAAGAGAGAGCGACTCTCAACTTTTATGATCCAA	257		
QY	2186	ATCCTCTTATGATTTTTCATTTTTTTGTGTGCTTATATGTCATGCCACCAATGATATGAAAA	2245		
Db	258	ACAATTTAATATGACATCGGGGTGCTGGCGATGATTAAGGCGATCCATACAGTAGATCTT	317		
QY	2246	CAGACTACGCAAAATTTCCCTATGTTGTAAGAAAACACCTGAGAGAAACATTTGTGGGTATAG	2305		
Db	318	GAGCTTGAGATATCTTCCCTCCGCAAAAGAACCACTGGAACATGAGAGTAGGCAACAG	377		
QY	2306	AGTGCAGAGGAGCATGCTCTATTTGATTTGATGAA	2342		
Db	378	GTTCGTGAGAAAGTATTCACCTCTCATGGCTTGA	414		
RESULT	12				
LOCUS	BG284445/c				
DEFINITION	602408575f1 NIH_MGC_91 Homo sapiens CDNA clone IMAGE:4520466 5',				
ACCESSION	BG284445	806 bp	EST	21-FEB-2001	
VERSION	BG284445.1	GI:13035400			
KEYWORDS	EST.				
SOURCE	human.				

REFERENCE	TITLE	AUTHORS	JOURNAL	COMMENT
ORGANISM	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	NIH-MGC	http://mgc.ncl.nih.gov/		
TITLE	1 (bases 1 to 806)			
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgaabs.r@email.nih.gov Tissue Procurement: DCTD/DTP CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMLN at: http://image.ln1.gov			
FEATURES	plate: LLM10418 row: f column: 19 High quality sequence stop: 711. Location/Qualifiers			
source	1. 806 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4520466" /clone_1ib="NIH-MGC-91" /russue_type="adenocarcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: prostate; Vector: pCMV-SPORT6, site_1: NotI; site_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 1.4 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."			
BASE COUNT	249 a 158 c 188 g 211 t			
ORIGIN				
Query Match	1.3%; Score 40.2; DB 175; Length 806; Best Local Similarity 54.4%; Pred. No. 5.8;			
Matches	81; Conservative 0; Mismatches 68; Indels 0; Gaps 0;			
OY	615 AGTAATCAGGCAGATGAATTTGGACAAATATCTGCCAAGAAATGTTCTGAGGGCC 674			
Db	712 ACTACTTTATGATACAGCTACACATACACAGAAAGTTACTGGAATGCTCGGATTAAGA 653			
OY	675 TTTGTACCCGATGGTATTTGATATGATGATGACAGAACTCTATGTTATTCGAGAAA 734			
Db	652 TTGTTTTCCTGTGTCATTTTGTCTTTTTPACAGAGTTTTCCTTCCTCGAGATTA 593			
OY	735 TGAGCATTTGACTACTCTATGATAGTAGG 763			
Db	592 TAATGAACATGTGTACACACACAGATTAAG 564			
RESULT 13				
CNS00B2A	1101 bp DNA GSS 04-JUN-1999			
LOCUS	Drosophila melanogaster genome survey sequence T7 end of BAC #			
DEFINITION	BACR22B23 of RPCR-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	AL056333			
VERSION	AL056333.1 GI:4937102			
KEYWORDS	GSS.			
SOURCE	fruit fly.			
ORGANISM	Drosophila melanogaster			
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
AUTHORS	1 (bases 1 to 1101)			
TITLE	Genoscope.			
JOURNAL	Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr - web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a			



collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

Location/Qualifiers

1.1101

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_lib="RPCI-98"

/note="BACR22B23"

/note="end : 17"

BASE COUNT 235 a 165 c 186 g 372 t 143 others

## ORIGIN

Query Match 1.3%; Score 40.2; DB 219; Length 1101;

Best Local Similarity 40.0%; Pred. No. 6.3;

Matches 80; Conservative 28; Mismatches 92; Indels 0; Gaps 0;

OY 280 TATTTTAAATATTTTCCTCAGAGAACTACAAGAGGAGGCTTACTTTGTCCAAAT 339

Db 693 WATTTTGCATTTKTAACACAGACAGAAATTAATTTGGATATTTTCATATAT 752

OY 340 CCTCAGCAAGGCGACGCTTTTCTGGGTTCTCCAGCCTCTTTTATTCAGACCCCGGA 399

Db 753 TTTCMACACGCGGCTTTTGTGATCCCTCCTCTTTTTRSCCTTCMMVSM 812

OY 400 GATATTAAGAAACGAGATCTCTATTCAAAATGACATTTATGCTTAAACATAT 459

Db 813 TTTTGGGGGCTCGGATTTTWTGTTTAAAGTCTTTTGTGNGKRAAAATTTA 872

OY 460 GTAGTGGCTTTTGACACAAA 479

Db 873 WTARTGATTTKKGRRARA 892

RESULT 14

A0768358 545 bp DNA GSS 28-JUL-1999

LOCUS HS\_3162.A1.G10.TTC CIT Approved Human Genomic Sperm Library D Homo

DEFINITION sapiens genomic clone Plate=3162 Col=19 Row=M, DNA sequence.

ACCESSION A0768358

VERSION A0768358.1 GI:5646474

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 545)

Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and

Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

CONTACT: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 3162 row: M column: 19

Seq primer: 17

Class: BAC ends

High quality sequence stop: 545.

Location/Qualifiers

1.545

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Plate=3162 Col=19 Row=M"

/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 177 a 101 c 86 g 168 t 13 others

ORIGIN

Query Match 1.3%; Score 40; DB 233; Length 545;

Best Local Similarity 55.9%; Pred. No. 5.9;

Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

OY 826 TCTGACATTAACAGTCTTGTGAAGAAACCATTCATATGGGTGGAGCCAT 885

Db 387 TCATACATTTTTCAGTAACTGATGATGATGATGATGATGATGATGATGATGAT 446

OY 886 TATGCTAGGAACCTTACATCTCTCAGAGGCTCTACTCTATTTATCAATATATCA 945

Db 447 TATGCTAGGAACCTTACATCTCTCAGAGGCTCTACTCTATTTATCAATATATCA 506

OY 946 TATGCAATTCGCAAA 961

Db 507 TATGCACTTAGGACA 522

RESULT 15

A0793076 563 bp DNA GSS 03-AUG-1999

LOCUS HS\_5305.A2.A05.T7A RPCI-11 Human Male BAC Library Homo sapiens

DEFINITION genomic clone Plate=881 Col=10 Row=A, DNA sequence.

ACCESSION A0793076

VERSION A0793076.1 GI:5700700

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 563)

Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and

Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

CONTACT: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm))

or from Resear h Genetics (info@resgen.com). BAC end Web Server:

<http://www.htsc.washington.edu>

Plate: 881 row: A column: 10

Seq primer: 17

Class: BAC ends

High quality sequence stop: 563.

## FEATURES

**Source**

Location/Qualifiers  
1563

1.563

```
/organism="Homo sapiens"
/dbxref="taxon:9606"
```

```
/db_xref="taxon:9606"
/colname="plate-991_col"
```

```
clone="Plate=881 Col=10 Row=A"
```

```
/clone_lib="RPC1-11 Human Male BAC Library"
```

```

/crone_11b-
/sex="male"

```

```
/sex="male";
/notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
```

Male blood DNA was isolated from one randomly chosen donor  
/note="Vector: pBACE3.6; site\_1: EcoRI; site\_2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of *EcoRI* and

and partially digested with a combination of *EcoRI* and *EcoRI* Methylase. Size selected DNA was cloned into the

ECORI methylase. Size selected

BASE COUNT	173 a	104 c	102 g	176 t	8 others
BASE COUNT	173 a	104 c	102 g	176 t	8 others

ORIGIN

Query Match  
Best Local

Best Local Similarity 59.8%; Pred. No. 6;  
Matches 67; Conservative 0; Mismatch

Matches 67; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

2/01 6GAGTATAGAAATGCCGCCCAAGCTAGGAATTAATAATAAACCTGTGGAGCAAAATT 2/60

Db 432 GTAGTCTGTAATGACACCAAGATTATAGTACGTCATGCATAGAGTGAAGTGAAGAAAGTC 373

Qy 2761 CGTTTTAGAGGTTTCATTGCCCTGTGTGTTCCGATCTTAATAAAT 2812

Db 372 AATTTTGTAACTCGCCATTGGCTTTTGTCTTTTGGCTTCACTTAAAT 321

Search completed: October 2, 2001, 03:06:34  
Job time: 51652 sec

Job time: 51652 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: October 2, 2001, 03:26:17 ; Search time 97.54 Seconds  
(without alignments)  
573.050 Million cell updates/sec

Title: US-09-446-677b-6

Perfect score: 4814  
Sequence: 1 MRFSLGGRPIVFSILTLVSF.....IECRHANNVINCGRFRF 922

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

A.Geneseq.0601:\*  
1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4814	100.0	922	20	AAW88419
2	4789	99.5	922	20	AAW88419
3	4769	99.1	922	20	AAW88419
4	2099	43.6	400	20	AAW88430
5	1467.5	30.5	918	21	AAW88430
6	1444	30.0	928	21	AAW88430
7	1442.5	30.0	918	20	AAW88430
8	1434	29.8	928	20	AAW88430
9	1430.5	29.7	928	21	AAW88430
10	1426	29.6	930	21	AAW88430
11	1426	29.6	930	21	AAW88430

12	1423.5	29.6	928	20	AAW88421	Chlamydia pneumoniae
13	1423	29.6	930	20	AAW88424	Chlamydia pneumoniae
14	1405.5	29.2	927	20	AAW88424	Chlamydia pneumoniae
15	1400.5	29.1	925	21	AAW88424	Chlamydia pneumoniae
16	1400.5	29.1	936	21	AAW88424	Chlamydia pneumoniae
17	1393	28.9	928	21	AAW88424	Chlamydia pneumoniae
18	1356	28.2	928	21	AAW88424	Chlamydia pneumoniae
19	1346	28.0	918	20	AAW88424	Chlamydia pneumoniae
20	1345	27.9	949	20	AAW88424	Chlamydia pneumoniae
21	1344	27.9	928	21	AAW88424	Chlamydia pneumoniae
22	1344	27.9	928	21	AAW88424	Chlamydia pneumoniae
23	1325.5	27.5	885	21	AAW88424	Chlamydia pneumoniae
24	1262	26.2	945	21	AAW88424	Chlamydia pneumoniae
25	1250	26.0	945	21	AAW88424	Chlamydia pneumoniae
26	1131.5	23.5	973	21	AAW88424	Chlamydia pneumoniae
27	1012	21.0	643	21	AAW88424	Chlamydia pneumoniae
28	1007	20.9	841	21	AAW88424	Chlamydia pneumoniae
29	996	20.7	841	21	AAW88424	Chlamydia pneumoniae
30	959	19.9	1006	21	AAW88424	Chlamydia pneumoniae
31	958	19.9	982	21	AAW88424	Chlamydia pneumoniae
32	957	19.9	1012	20	AAW88424	Chlamydia pneumoniae
33	948.5	19.7	1013	20	AAW88424	Chlamydia pneumoniae
34	941.5	19.6	671	20	AAW88424	Chlamydia pneumoniae
35	931.5	19.3	1013	20	AAW88424	Chlamydia pneumoniae
36	872	18.1	597	20	AAW88424	Chlamydia pneumoniae
37	868.5	18.0	507	20	AAW88424	Chlamydia pneumoniae
38	832.5	17.3	1132	20	AAW88424	Chlamydia pneumoniae
39	747	15.5	880	21	AAW88424	Chlamydia pneumoniae
40	746	15.5	860	21	AAW88424	Chlamydia pneumoniae
41	734	15.2	866	21	AAW88424	Chlamydia pneumoniae
42	716.5	14.9	530	20	AAW88424	Chlamydia pneumoniae
43	686	14.3	610	20	AAW88424	Chlamydia pneumoniae
44	611	12.7	1146	20	AAW88424	Chlamydia pneumoniae
45	609.5	12.7	483	20	AAW88424	Chlamydia pneumoniae

## ALIGNMENTS

RESULT	1	
ID	AAW88419	standard; Protein: 922 AA.
XX	AAW88419:	
AC	26-APR-1999	(first entry)
XX		
DE	Chlamydia pneumoniae surface exposed protein Omp6.	
KW	Omp6; outer membrane protein 6; surface exposed protein; antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.	
XX		
OS	Chlamydia pneumoniae.	
XX		
PN	W09858953-A2.	
XX		
PD	30-DEC-1998.	
XX		
PF	19-JUN-1998:	98MO-DK00266.
XX		
PR	23-JUN-1997:	97DK-0000744.
XX		
PA	(BIRK/) BIRKELUND S.	
PA	(CHRI/) CHRISTIANSEN G.	
PI	Birkelund S.	Christiansen G.
PI	Myrdal P.	Knudsen K.
XX		
DR	WPI: 1999-105610/09.	Madsen A.
DR	N-PSDB; AAX06818.	
XX		
PT	Species-specific test for identifying mammals infected with Chlamydia pneumoniae - comprises detecting antibodies specific for	

PT outer membrane proteins of *C. pneumoniae* or nucleic acids encoding  
 PT these proteins  
 PS Claim 7: Page 47-49; 115pp; English.  
 XX  
 CC This polypeptide comprises the novel 100.3 kDa surface exposed  
 CC protein Omp6 of the human respiratory pathogen *Chlamydia*  
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see  
 CC AA006818) isolated from a *C. pneumoniae* expression library. The  
 CC invention provides 12 novel surface exposed proteins, Omp4-Omp12  
 CC (see AA06817-28), and nucleic acid sequences encoding them (see  
 CC AA06816-27). A new species specific test is claimed that is used  
 CC to identify mammals (including humans) infected with *Chlamydia*  
 CC pneumoniae. The test comprises detecting antibodies specific for  
 CC Omp4-Omp12 or detecting nucleic acid fragments encoding these outer  
 CC membrane proteins, especially by PCR. The proteins are also used  
 CC in the diagnosis of *C. pneumoniae* infection in mammals. The  
 CC nucleic acids and proteins can also be used in the immunization of  
 CC mammals, the nucleic acids being particularly useful as DNA  
 CC vaccines for eliciting in vivo expression of antigens. The  
 CC vaccines may also prevent atherosclerosis and bronchial asthma,  
 CC which are possibly associated with *C. pneumoniae*.  
 CC  
 XX Sequence 922 AA:  
 SQ  
 Query Match 100.0%; Score 4814; DB 20; Length 922;  
 Best local Similarity 100.0%; Pred. No. 0;  
 Matches 922; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 MRSLSGPFVLESLLSLVDTLSATTSILPDESHGDSQNAENSYNOAGDVSLTG 60  
 DB 1 mrfslgfpvlvfltlsvrfdtslatsltpedsfhgdsqnaersynvqadvyaltg 60  
 OY 61 DVISVNDNALNKAFCFNTVSGSVTFAGNHGILYFNINISGTRKEGVILCCDPQATAR 120  
 DB 61 dvsvndnalknacfntvsgsvtfagnhgilymnissgtrkegvilccdpqatart 120  
 OY 121 SGFSTLSFTOSPDIREQGLYKRNALMLNNTVVRFEQNSKTKGALSGAVNTVGN 180  
 DB 121 sgfstlsftospdireqglykernalmlnntvvrfeqnsktkgalsganvltvgny 180  
 OY 121 sgfstlsftospdireqglykernalmlnntvvrfeqnsktkgalsganvltvgny 180  
 DB 121 sgfstlsftospdireqglykernalmlnntvvrfeqnsktkgalsganvltvgny 180  
 OY 181 DVSFFQONATFEGAIHSSGPILOIAVNOAETRAQNTAKGSGALYSDGDDIDONAVY 240  
 DB 181 dvsffqonatifegaihssgpiioiavnnoaetraqntakgsgalysdgdiddidnnavy 240  
 OY 241 LPEFNEALTTAIGKGAACVCLPTSGSSTPVPIYTFSONKOLVPERHNSIMGGAITARKL 300  
 DB 241 lfenealttaigkgavccclptsgsstpvpiytfsonkolvperhnsimggaityarkl 300  
 OY 301 SLSGGPTLEFINNISVANSQNLGAIADTGRISLSAEKGIITFGQNRSLPFLNGIHL 360  
 DB 301 slsggptlefinnisvansqnlgaiaidtgrislsekgiitfgqnrtslpflngihl 360  
 OY 301 slsggptlefinnisvansqnlgaiaidtgrislsekgiitfgqnrtslpflngihl 360  
 DB 301 slsggptlefinnisvansqnlgaiaidtgrislsekgiitfgqnrtslpflngihl 360  
 OY 361 LONAKRLKLOARNGCSIEPTDPTSEADSGTOLNINDEPKNKRYTGTILFSGKSLANDP 420  
 DB 361 lgnakrlkloarngcsieftdptseadsgtqlnindpknkrytgtilfsgkslandp 420  
 OY 421 RDEKSTIPQNVNLSAGVLYKEGAETVSKFQSPGSHVLDLGTLLASKEDIAITGTA 480  
 DB 421 rdkstipqnvnlsgvlykegaetvskfqpshvldlgtllaskediaitgta 480  
 OY 481 IDISLSSSSTAIVIKANTANKOISVTSIELISPTGNVYEDLRMNSQTEPPLSLPEGA 540  
 DB 481 idslssstaivikantankoisvtsielisptgnvyedlrmsqtepplslpega 540  
 OY 541 GGSVYTVAGDGLFVSHYHGQNMKIAMTGTGKKGVEFFEDKATINYRPREKESNLVNL 600  
 DB 541 ggsvtytvagdlfvsyhyhgqnmkiamtgtgkkgveffedkatinyrprekesnlvnl 600  
 OY 601 WGAIVAVRSIMOVQETHASLOTDRGLMTDIGNFHVASSENINRYRHNSGGVYLSVN 660  
 DB 601 wgaivavrsimovqethaslotdrglmtdigntfhuvasenninryrhnsggvylsvn 660  
 DB 601 wgaivavrsimovqethaslotdrglmtdigntfhuvasenninryrhnsggvylsvn 660

OY 661 ETPKHYTSMASFQSFSDKDYAVSNNEYKGLSYLYOTYSIGNIFRYASRNPVNW 720  
 DB 661 etpkhytsmasfqsfsdkdyavsnneykglsllyotysignifryasrnpvnw 720  
 OY 721 ILSRRLQNPMLTFHFLCAYGATNDKTDYANFPVKNRNNCAIECGGSPPLVYE 780  
 DB 721 ilsrflqnpmltfhflcaygatndktidyafpvknrnncaiecggspllyve 780  
 OY 781 NGRFQGAIRFEMLOLVYAVQGFKEETADGRFNSGSLTISVPLGIRPEKALASDYL 840  
 DB 781 ngrfogaifemlolvayvqgfketadgrfnsngsltisvplgirkpalkasdy 840  
 OY 841 YDFSFSYIPDIFKRDSCFCAALVTSQDMLVPAHVSRAVSGTGRVHFNDYTELLR 900  
 DB 841 ydfsfsyipdifkrdsccaalvtsqdmvpaahvsravsgtgrvhfndyellr 900  
 OY 901 GSIECRPHARNINCGSKRF 922  
 DB 901 gsiecrpharnincgskrf 922  
 DB 901 gsiecrpharnincgskrf 922  
 RESULT 2  
 ID AAY95548 standard; Protein; 922 AA.  
 AC AAY95548;  
 DT 10-OCT-2000 (first entry)  
 DE Chlamydia pneumoniae antigen CPN100877 RY-61.  
 DE CPN100877 RY-61; antigen; infection; diagnosis; therapy; vaccine;  
 KW outer membrane protein.  
 OS Chlamydia pneumoniae.  
 PN WO200039158-A1.  
 PD 06-JUL-2000.  
 PF 23-DEC-1999; 99WO-CA01230.  
 PR 23-DEC-1998; 98US-0113280.  
 PR 23-DEC-1998; 98US-0113281.  
 PR 23-DEC-1998; 98US-0113282.  
 PR 23-DEC-1998; 98US-0113283.  
 PR 23-DEC-1998; 98US-0113284.  
 PR 23-DEC-1998; 98US-0113285.  
 PR 28-DEC-1998; 98US-0114050.  
 PR 28-DEC-1998; 98US-0114056.  
 PR 28-DEC-1998; 98US-0114057.  
 PR 28-DEC-1998; 98US-0114058.  
 PR 28-DEC-1998; 98US-0114059.  
 PR 28-DEC-1998; 98US-0114061.  
 PA (CONN-) CONNAUGHT LAB LTD.  
 PI Murdin AD, Oomen RP, Wang J;  
 DR WPI: 2000-452369/39.  
 DR N-PSDB: AAA50035.  
 PT Novel Chlamydia polynucleotides and polypeptides useful for diagnosis,  
 PT prevention and treatment of Chlamydia infection in mammals -  
 PS Claim 16(a); Page 65-69; 215pp; English.  
 CC The present sequence is that of Chlamydia pneumoniae antigenic  
 CC protein CPN100877 RY-61, a putative 98 kDa outer membrane protein.  
 CC It is an example of *C. pneumoniae* antigenic polypeptides of the  
 CC invention (see AAY95543-55) that are encoded by open reading frames

CC (see AAY50030-42) identified in the *C. pneumoniae* genome. The  
 CC polypeptides are useful in the diagnosis, treatment and prevention  
 CC of Chlamydia infection. They can be prepared by recombinant  
 CC methods using transformed unicellular host cells. The  
 CC polypeptides, or immunogenic fragments of them, or fusion proteins  
 CC that include an additional polypeptide such as a heterologous  
 CC signal peptide or a polypeptide having adjuvant activity, are  
 CC utilised as vaccines to protect against Chlamydia infection.  
 CC The polypeptides are also used in diagnostic kits and in methods  
 CC of detecting Chlamydia infection.

CC  
 XX  
 SQ Sequence 922 AA:

Query Match 99.5%; Score 4789; DB 21; Length 922;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 918; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MRSFLGFLVFLSLTLLSVFDTSLSATTISLTPEDSPFHGDSQNAERSYVNOAGDVSLTG 60  
 DB 1 mrsflgflvflsltlsvfdtslsattisltpe dsfhgdsqnaersynvnoagdvsltg 60  
 QY 61 DVSISVNDNALNKKACFNVTSGSVTFAGNHGGLYFNINISSGTTKEGAVLCCQDPQATARF 120  
 DB 61 dvisvndnalnkkacfnvts gsvtfagnhgglyfninissgttkegavlocqdpqatarf 120  
 QY 121 SGFSTLSFIOSPEDIKEGGLYSKNMLMLNNTVVRFEQNSTKGAGISGAVNTIVGN 180  
 DB 121 sgfstlsfi ospedikeggl ysknmlmlnntvvrfeqnstk gagisgavntivgn 180  
 QY 181 DSVSFQNAATFGGAIHSSGPIQIAVNOARIFRQNTARKSGGALISPDIDIDONAVY 240  
 DB 181 dsvsfqnaatfggaihssgpi qia vnoarifrqn tarksggalis pdididonavy 240  
 QY 241 LFRNEALTTAIGKGA VCCIP TSGSSTPVPIYTFSDNKOLVERNHSLINGGAIYARKL 300  
 DB 241 lf rneal ttaigkga vccip tsgss tppvpiy tfsdnkol vernhsl inggai yarkl 300  
 QY 301 SSSGGPTLFINNISTANSONLGAIAIDTGGELSLSAEKGTTFGQNTSLPPLNGIHL 360  
 DB 301 sssggptlfinnistans onlgaia idtggel slsae kgttfgqntsl pplngihl 360  
 QY 361 LQNAKFLKLOARNGCSEIEFDPTSEADGSTOLINGDPPKNEYGTGLIFSGSKSLANDP 420  
 DB 361 lqnakflkloarngcseie fdp tseadgstol ingdppkneygt glifsgsksl andp 420  
 QY 421 RDRFKSTIPQNVNLASGLYVKEGAETVSKFTQSPGSHVLDTGLTKLIASKEDIAITGLA 480  
 DB 421 rdrfkstipqnvnl asg lyvkegaetvskftq spgshvldt gltkliaskedia itgl a 480  
 QY 481 IDIDSLSSSTAVITKANTANKOISTDSELSPTGNATEDLRMNSQTFPLLSERGA 540  
 DB 481 ididslssstavitk antank oistd sel sptgnated lrmnsq tffllsle rga 540  
 QY 541 GGSVTVAGDFLVPVSHYFGQGNMKLAWTGTGNKVGFEFMDKINYPKPEKENVLPNII 600  
 DB 541 ggsvtvagdf lvpvshy fgqgnmk l awtgtgnkv gfe fmdk inypkpek envlpn ii 600  
 QY 601 WGNNAVNRSLMOVETTHASSLQTRGLMIDIGNFTFVHSA SEDNIRYRHNSGGYVLSVN 660  
 DB 601 wgnnavnrslmo vetth asslqtr glm idignft fvh sasedn iryrhns ggyvls vn 660  
 QY 661 ETRPKHTYSAEQLSRQKDYAVSNNEKRYMIGSYLYQTTSLSGNTFRKASRNPVNVG 720  
 DB 661 etrpkytsaeq lsrqkdy avsnnek rymigsy lyq t t s l s g n t f r k a s r n p v n v g 720  
 QY 721 ILRRFLQNLMLFHFELCA GHA TNDKTDYANFPWKNSWRNMCATIECGSGMPLLYFE 780  
 DB 721 ilrrflqnlml fhfelc agha tndk tdyanf pwkns wrnmc a tiec gsgm plly fe 780  
 QY 781 NGRFLQGAIPFMKLIQLVAYAGDFKETTADGRFSSNGSLTSSVPLGIRREKALASQDVL 840  
 DB 781 ngrflqgaipf mkl iqlvay agdfkett adgrf s sngsl tssvpl g irrek al asq dvl 840

QY 841 YDSESYTIDIFRKPDSCEALYISGDSMLVPAHYSRHAFVSGTGRHFNNDYTELLCR 900  
 DB 841 ydsesytipd ifrkpdsce al yisgds mlvpa hy srh afvsgtgr hfnndy tellcr 900  
 QY 901 GSTECRPHANYNINGSKPRF 922  
 DB 901 gstecrph anynin gskprf 922

RESULT 3

AAY34597  
 ID AAY34597 standard; protein; 922 AA.

AC AAY34597;

DT 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae surface exposed polypeptide.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN W09927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98MO-1B01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97PR-0014673.

XX (GENSET) GENSET.

PA Griffais R;

PI WPI; 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

PS Page 620-621; Disclosure: 1912pp; English.

XX AAY34584-Y35879 represent the proteins encoded by all the open reading

CC frames in the complete genome (see AAY91990) of Chlamydia pneumoniae.

CC C. pneumoniae causes respiratory disease such as pneumonia and

CC bronchitis and is thought to be a contributing factor in heart

CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema

CC nodosum or pharyngitis. The polypeptides encoded by the open reading

CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in

CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae

CC nucleotide sequences can also be used as immunogenic compositions,

CC especially where the vector directs the expression of a neutralising

CC epitope of C. pneumoniae.

XX  
 SQ Sequence 922 AA:

Query Match 99.1%; Score 4769; DB 20; Length 922;  
 Best Local Similarity 99.2%; Pred. No. 0;  
 Matches 915; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MRSFLGFLVFLSLTLLSVFDTSLSATTISLTPEDSPFHGDSQNAERSYVNOAGDVSLTG 60  
 DB 1 mrsflgflvflsltlsvfdtslsattisltpe dsfhgdsqnaersynvnoagdvsltg 60  
 QY 61 DVSISVNDNALNKKACFNVTSGSVTFAGNHGGLYFNINISSGTTKEGAVLCCQDPQATARF 120  
 DB 61 dvisvndnalnkkacfnvts gsvtfagnhgglyfninissgttkegavlocqdpqatarf 120  
 QY 121 SGFSTLSFIOSPEDIKEGGLYSKNMLMLNNTVVRFEQNSTKGAGISGAVNTIVGN 180

```

Db 121 sgftslfngspdqikeggqlylsknaImInnyvrfeqgsktkgalsganvltvgny 180
OY 181 DSVSEFQNMATFEGAIHSSGPLOIAVNOAEIRFAQNTAKSGSGALYSQSDIDIDONAVY 240
Db 181 dsvsfyqnaatfsgaIhssgpIqIavngaeIrfaqnIakngsgaIysdIdIdqnavy 240
OY 241 LFEENALATTAIGKGAVCCLPTSGSSTPVPYITFSDNKLVEERHNSIMGGAITYARKL 300
Db 241 lfrenaalItaIkggavccIptsgsstpyrIvtfsdnkqlvferhnsImnggaIyarkl 300
OY 301 SSSGGPPLFINNISSANSONLGAIALDTGGEISLSAEKGTTFGQNTSLPFLNGIHL 360
Db 301 sIsaggpIlfInnisysanqnIlgalaIdtgeIsIsaekgtIltfgnrtslpflngIhl 360
OY 361 LQNAKELKIQARNGCGSEEDPITSEADSTONINGDPNKKREYTGIIIFESGKSILANDP 420
Db 361 lqnakelIkIqarngysIeIdpItseadsgtqlInngdpknkeytqIIfsgeksIandp 420
OY 421 RDRFSTIPQNVNLSAGYLVIKEGAETVSKFTQSPGSHLVLDGTGLIASKEDIAITGLA 480
Db 421 rdfkstIppqnvnIsagylvIkegaetvskftqspshlvldgtgLIaskedIaltglA 480
OY 481 IDIDSJSSSTAAVITAKNTANKOISYTDSTIELISPTGNAYEDILMRNSQTFPLLSBPQA 540
Db 481 IdIdssstaaVItakntankqIsytdstIeIIsptgnayEdIlmrnsqtFpllsbpqa 540
OY 541 GGSVYTWAGDFLVPSPHYGFGQNMKLAWTGTGNKVGEEFFMDKINTYPRPEKEGNLVNII 600
Db 541 ggsvtwtagdfIvpshpyfgqnmkIawtgtnkvgeffmdkInyprpekegnlvnII 600
OY 601 WGNNAVNRSLMOQVETTHASSLQTDRLGWLIDIGLGNFHVSASEDNIRYRNHSGGYLSVNN 660
Db 601 wgnnavnrslmoqVetthasslQtdrLgwlIdIGlgnFhvsaSednIryrnhsggyLsvnn 660
OY 661 EITPKRYTSMASFQSLSRDKDYAVSNNEERMYIGSYLYQTTSLGNIFFRASNPNVNWG 720
Db 661 eItpkrytsmasfQslsrDkdyavsnneermyIgsylyqTtSLgNIffrasnPNvNWg 720
OY 721 IISRRFLQNPMLFHFLECAAGHATNDMKTDYANFPMVKNSMRNCAIECGSGMPLLVFE 780
Db 721 IIsrrflQnpmlfHfLeCaagHAtndMktDyanfpmvknsmrncAIEcgsgmPlLVfe 780
OY 781 NGRLEFGALPFRMLQLVAYAGDFKETTADGRFNSGSLTSISVPLGIRREKLALSQDVL 840
Db 781 ngrlEfgalPfrmlQlvayAGdfkettADgrfnsGsltsIsvplGirreklalsqdvl 840
OY 841 YDSEFSYIPIPFKDPSCERALVYISGDSMLVPAHVSRAHAFVSGRGHFNENTYELTLCR 900
Db 841 ydSefsYIpiPFkdpScEralvYIsGdsmlVpAhvsrAhafvSGrgHfnEntyELtLCr 900
OY 901 GSIECRPHARNYININGSKFRF 922
Db 901 gsIecrpharnyInIngskfrf 922
OY 901 gStecrpharnyInIngskfrf 922

```

```

RESULT 4
AAW88430
ID AAW88430 standard; Protein; 400 AA.
XX
AC AAW88430;
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein omp6.
XX
KM Omp6; outer membrane protein 6; surface exposed protein; antigen;
XX Infection; diagnosis; vaccine; atherosclerosis; asthma.
OS Chlamydia pneumoniae.
XX
PN W09858953-A2.
XX

```

```

PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98W0-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Myrind P;
XX
XX WPI: 1999-105610/09.
DR N-PSDB; AAX06829.
XX
PT Species-specific test for identifying mammals infected with
PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT these proteins
XX
PS Disclosure: Page 82-83; 115pp; English.
XX
CC This polypeptide is described as a subsequence of the claimed
CC novel surface exposed protein Omp6 (see AAW88419) of Chlamydia
CC pneumoniae, a human respiratory pathogen. The invention provides
CC a new species specific test for identifying mammals (including
CC humans) infected with C. pneumoniae. The test comprises detecting
CC antibodies specific for surface exposed proteins Omp4-Omp15 (see
CC AAW88417-28) or detecting nucleic acid fragments encoding them (see
CC AAX06816-27), especially by PCR. The proteins are also used in the
CC diagnosis of C. pneumoniae infection in mammals. The nucleic acids
CC and proteins can also be used in the immunization of mammals, the
CC nucleic acids being particularly useful as DNA vaccines for
CC effecting in vivo expression of antigens. The vaccines may also
CC prevent atherosclerosis and bronchial asthma, which are possibly
CC associated with C. pneumoniae.
XX
XX Sequence 400 AA:
SQ

```

```

Query Match 43.6%; Score 2099; DB 20; Length 400;
Best Local Similarity 99.5%; Pred. No. 4, 2e-157;
Matches 398; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 398 DPNKREYTGIIIFSGEKSILANDPRDFKSTIPQNVNLSAGYLVIKEGAETVSKFTQSPGS 457
Db 1 dpnkreytgIIIfsgeksIandprdfkstIppqnvnIsagylvIkegaetvskftqspgs 60
OY 458 HLVLDTGTLKILASKEDIAITGLAIDISLSSSTAAVITAKNTANKOISYTDSTIELISPTG 517
Db 61 hlvldgtkLIaskedIaltglAIdIdssstaaVItakntankqIsytdstIeIIsptg 120
OY 518 NAYEDILMRNSQTFPLLSBPAGGSVYTAGDFLVPSPHYGFGQNMKLAWTGTGNKVGEE 577
Db 121 nayedIlmrnsqtFpllsbpAggsvYtagdfIvpshpyfgqnmkIawtgtnkvge 180
OY 578 FFWDKINTYPRPEKEGNLVNIIIMGNAVNRSLMOQVETTHASSLQTDRLGWLIDIGLGNF 637
Db 181 ffwdkInyprpekegnlvnIIwgnnavnrslmoqVetthasslQtdrLgwlIdIGlgnfh 240
OY 638 VSASEDNIRYRNHSGGYLSVNNETPKHTYSMAFSQLSRDKDYAVSNNEERMYIGSYL 697
Db 241 vsasEdnIryrnhsggyLsvnnetpkhtysmafsqLsrDkdyavsnneeymyIgsyl 300
OY 698 YQYTTSLGNIFFRASNPNVNWGIIISRRFLQNPMLFHFLECAAGHATNDMKTDYANFPMV 757
Db 301 yqytslgnIffrasnPNvNWgIIIsrrflQnpmlfHfLeCaagHAtndMktDyanfpmv 360
OY 758 KNSWRNNCWAIECGSGMPLLVFENGRLFOGAIPFMKIQLV 797
Db 361 knswrncwaIkcgsmPlLVfengklfGgalpfmklqIv 400

```





PD	08-JUN-2000.
XX	
PF	01-DEC-1999; 99WO-CA01147.
XX	
PR	01-DEC-1998; 98US-0110339.
PR	01-DEC-1998; 98US-0110340.
PR	01-DEC-1998; 98US-0110427.
PR	01-DEC-1998; 98US-0110428.
PR	01-DEC-1998; 98US-0110438.
XX	
PA	(CONN-) CONNAUGHT LAB LTD.
XX	
PI	Murdin AD, Oomen RP, Wang J;
XX	
DR	WPI; 2000-412339/35.
DR	N-PSDB; AAA30851, AAA30852.
XX	
PT	Nucleic acids encoding polypeptide antigens from Chlamydia useful for
PT	preventing, diagnosing and treating diseases such as community acquired
PT	pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
XX	asthma -
XX	
DS	Claim 16; Fig 5; 174pp; English.

This sentence is a Chlamydia antigen of the invention, designated CC CN100638. The nucleic acids (and their complementary sequences) may be used as diagnostic agents for detecting the presence of nucleic acids encoding Chlamydia antigens in samples according to standard methods, and therefore, for diagnosing Chlamydia infections. For example, they may be used as primers and probes for diagnostic polymerase chain reaction (PCR) assays. Antisense sequences may be used to down regulate expression of the proteins and may be used to treat infections. The nucleic acids may also be used to produce the protein antigens they encode according to standard recombinant DNA methodologies. The proteins may then be used as antigens for the production of antibodies (i.e. as vaccines) for preventing infection by Chlamydia. The antibodies may also be used as diagnostic reagents for detecting infections. Chlamydia is a pathogen implicated in the development of (for example) community acquired pneumonia, upper respiratory tract disease (especially bronchitis and sinusitis), asthmatic bronchitis, adult-onset asthma and acute exacerbations of asthma in adults.

50 Sequence 928 AA;

Query Match	30.0%;	Score 1444;	DB 21;	Length 928;
Best Local Similarity	36.2%;	Pred. No. 6.8e+105;		
Matches 348;	Conservative 168;	Mismatches 374;	Indels 70;	Gaps 25

Oy	1	MRSILCGFPLVSLTLV-LSV-FDRLSMTWTLSLPEDSFGSDGNAERXYNQA-GDYVS	57
Dd	1	mkslhwflsiisslalplslinsfsaaavealnlpctnsfsgpylrppactnagdllyln	60
Oy	58	LNGDVSIWNVDN-SALNKACEFNVTGSVTEFAGNHGLYFNINISGTETKEAGVALCCDDPQA	114
Dd	61	ltgdvsltnagaptiltascfcfettgnysfgbhqyqlfnldagan-----cftnt	114
Oy	117	TAR----FEGSGFTLFIOSPDIDKEOCIVSKNMMLNNVVYRFQONSKTKGASIGA	177
Dd	114	aankllisfgfysljqlctnatgqyalstgacsigsnyscyfignsndngalags	177
Oy	173	NWTVGNVDVSFYTONAAT-EGAIHSSGPLQIAVNQAEIRFAONTAKKSGGALSDD	233
Dd	174	srlslsln-pultfarnkatqkgalystgftlnnltnasfsentean-ngalayteas	233
Oy	232	IDIDONAAYLFRENALTTALGGAACCLPTSGSPRYPIVFSNOKLOPERHNHSIG	293
Dd	232	sflsnkalsf-inmsvataatggalye---sstepkvvlslsdngelnfigntalts	287
Oy	292	GGAIYARKLTISSGGPTLFINNISTANSOULGALAIDTGETISLAERKGTTFPGN---	344
Dd	288	gaailydcmlnvlsagpplflknmsgydaapgyalaadgsalslaaygdltfgncrv	344

OY	349	-----RTSLPFLNGIHLONAFLKLOANCSIEFYPIRTE--ADGSTOLIN	396
Db	348	kgaassqcttnsinlign-----tnakivqlraagqntlfyfydobltsitaalsda	402
OY	397	GD--PKNKEYTGLTFEGEK---SLANDPRFKSTIPQONVNLASGYLVIKEGA	451
Db	403	gpdlaognpaygltlivesgekliseaeaaadnlkibqlptltaggqslksyvtl	462
OY	452	TQSGSHLVLDLGRKLASKEDNIAITGLADIDLSSSSTPAVIKANTAKOISYD	511
Db	463	sqspgstlmaagctll-etadgltlmlvlnvdsiket-kgtlkatqasqvltlsg	520
OY	512	LISPTGNAYEEDLRMRNSOTPFLFLEPCAGAGSYVT--AGDPFLVSP-HG	568
Db	521	lvdpagnvvedvsnmpvyscltltadcpnrlhtltaadplekpnlnhgvygnal	560
OY	569	---TGCGNKVEEFEMDXINXKPREKEGKLVPLNLMGNAVNSLMQVOETHAS	625
Db	581	gedatatskaatlctwtktgynpberqgtlvantlwsfvdivsldqtlvakt	640
OY	626	GLMWIDIGNEFHVASBEDNIRYRNSGCVLVSVNNEITPKHYTSMASFQ	685
Db	641	glwcegslnffhkdstklnkfrrlnsagvygattlasdnlltaacqjfygdr	700
OY	666	NNETRMVIGSLYQY--TTSIGNIFRKASRNPVNVNGISLRRLQNPMLF	743
Db	701	knrasayaaslhldhlatlaspalllpsses-----eqp-vlfdqayl	748
OY	744	TNDKKTVDANPWNKNSMRBNKCAIEECGSNPLFVFNCRGLFOGADP	803
Db	749	knmktytqarbkesswynogcalaelasrlptalsheglfinafryflkve	808
OY	804	FKE--TTADGRFRSSGLTSSIVPLGIRFEKLASQDVLVDFSFYSIPDI	862
Db	809	fkenrtlvtsfidsgdllnvsyprgltfetrfsmerasayeatlvuadvy	868
OY	863	VISGDSWLVPAHNSRIAPFGSGGGRHNFDYTLFLLCRGSIETRPHARN	922
Db	869	lmlntswkvtcgtlnsrqagryraqrlfyaflpnlevtslnsmelstgssr	928

## RESULT 7

ID AAW88422 standard; Protein; 918 AA.

AC AAW88422;

DT 26-APR-1999 (first entry)

DE Chlamydia pneumoniae surface exposed protein Omp9.

KW **Omp9**; outer membrane protein 9; surface exposed protein; antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.

OS Chlamydia pneumoniae.

PN W09858953-A2

PD 30-DEC-1998.

PF 19-JUN-1998; 98WO-DK00266.

PR 23-JUN-1997; 97DK-0000744.

PA (BIRK/) BIRKELUND S.  
PA (CHRI/) CHRISTIANSEN G.

XX Birkelund S, Christiansen G, Knudsen K, Madsen A, Pi

PI Mygind P;  
.....

DR WPI; 1999-105610/09.  
DR N-PSDB; AAX06821.

XX

PT Species-specific test for identifying mammals infected with  
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for  
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding  
 PT these proteins

PS Claim 7; Page 56-58; 115pp; English.

XX This polypeptide comprises the novel 96.7 kDa surface exposed  
 CC protein Omp5 of the human respiratory pathogen Chlamydia  
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see  
 CC AAX06821) isolated from a C. pneumoniae expression library. The  
 CC invention provides 12 novel surface exposed proteins, Omp4-Omp15  
 CC (see AAW88417-28), and nucleic acid sequences encoding them (see  
 CC AAX06816-27). A new species specific test is claimed that is used  
 CC to identify mammals (including humans) infected with Chlamydia  
 CC pneumoniae. The test comprises detecting antibodies specific for  
 CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer  
 CC membrane proteins, especially by PCR. The proteins are also used  
 CC in the diagnosis of C. pneumoniae infection in mammals. The  
 CC nucleic acids and proteins can also be used in the immunization of  
 CC mammals, the nucleic acids being particularly useful as DNA  
 CC vaccines for effecting in vivo expression of antigens. The  
 CC vaccines may also prevent atherosclerosis and bronchial asthma,  
 CC which are possibly associated with C. pneumoniae.

SO Sequence 918 AA;

Query Match 30.0%; Score 1442.5; DB 20; Length 918;

Best Local Similarity 36.3%; Pred. No. 8.7e-105;

Matches 346; Conservative 152; Mismatches 375; Indels 79; Gaps 23;

QY 12 FSLTSL-----VDTSLSATITSLTPDSFHGDSQNAE---RSYVQAGDVSLTG 60  
 DB 5 fslslslslafhlsvsdadaadltsrdsyngdstetfepkaatsdagstlyldg 64  
 QY 61 DVASISNV-DNSAINKKACENTSGSVMPAGNHHLTFENISGTTKEAVLCCDDPQATAR 119  
 DB 65 dvasisngqktsltscfscfnatagnltlgnfshdnllstlv-agvvsnntaasgik 123  
 QY 120 FSGFSTLSFIOFSGDIREGCLYSKNALMLNMYVVE-----QNSKTRKGALSG 171  
 DB 124 fsgfstlmlaaprtt-----gkgaikltdglv--fsglndqemnsengsgaint 174  
 QY 172 ANVTYGNIDSVFQYONAAI-FGCAIHSGLPQIAVNOAIFRAONTAKKSGGALYSDG 230  
 DB 175 ktlstfgstrfvalfnssgqgaiaysagdvysenagllsfngnsalt-sggaalsag 233  
 QY 231 DIDIDQNAVYLFRENEALTAIGKGGAVCCLPRTSGSTPYPIYFSPNKLVERNSIM 290  
 DB 234 nvlsmgnqiffigckattc---nggalidc--nkaganpdlitlsqnesihlfnlnagn 287  
 QY 291 GGGAIYARKLISISSG-GPTLFINNISYANSQNLGAIADTGEISLSAEKGTTFQGNR 349  
 DB 288 sggaikytkkivlasggyvifsmn-kaanaetpggalaiaidsgelsiadlgnlffegnt 346  
 QY 350 TSLP-----FLNGIHLQAKFLKIQARNGCSIEFYDPTTSE-ADGSDTLANIGDPKMK 402  
 DB 347 tltstspasvnlraiaidnasnaklfnlratrgnkvlfydrptltsagatcklslnkadsgsn 406  
 QY 403 EYTGTLFSGSEKSLAND---PRDFKSTIPONVNLISAGYLIRKGAETVVKFQSPSHL 459  
 DB 407 tyegylivfsgseklseelkklkklkltfcgavelaagalyikdgvevvanlticqveskv 466  
 QY 460 VLDLGLTKLASKEDIAITGLAIDISLSSSTAIVAKANTANKQISVTDSEILISPTGNA 519  
 DB 467 vmdggttfesaeagvclngjalndisldgtn-kalkataaskdvalsgrlmvldaguny 525  
 QY 520 YEDLRMNSTPFLSLSEFAGGSVTVTAGDFLPVSP-----HYGQGWKMLAWGTG- 572  
 DB 526 yehlnlsqgqvlfliel--saqgtmttcd---ipdclplntnthhyqgvgclllvwwdat 580  
 QY 573 --NKVGEFFFDKINYKRPKEKGNLVNIIWGNAVNVRSLMQOETRHASSLQIDRGLMID 630

DB 581 akknatlwtcktygknpnpergplvpnslwgsfvdvrsiqslmdstslsstlnlws 640  
 QY 631 GIGNFHVASSEDNIRYRNSGGVLSVNNETPKHAYTSAFSQLSRDKQDAYANNEYR 690  
 DB 641 giadflhedqkngqrsyrsasagyalgggfftasentffafacqlfygdkdhlvaknht 700  
 QY 691 MYLGSYLYQYTTSLIGNFRASRNPNVNCILSRFLQNLMLFHLCAAGHATNDMKT 750  
 DB 701 vyagamsyrrlsgesktlakllsgnsdslpfvfnarf-----ayghdtnmtlk 748  
 QY 751 YANFPWKNRNNCMAIECGSMPLLVFENGRLFGAIFPMKLQLYAVAGDFKETAD 810  
 DB 749 yltgspyksgwgnadafgiecgaipvvaasgrswdthpflnemyahqndtkengte 808  
 QY 811 GRFNSGSLTSISVPLGIREFKALSDVLYDPSFYIPIDFRKDPSCALYISGDSWL 870  
 DB 809 gsfqsedlfnlavpygikfek--fsdkstydlslayvpdvirndpcttlmwsgdsws 866  
 QY 871 VPAHYSRAHFVSGTGRVHFNDYTELLCGSIFCRPHANNINCGSKTRF 922  
 DB 867 tclgslsrqalivragnhlaasnfefvsgflevelygrsrsyaldigrirgf 918

# RESULT 8

AAW88423 ID AAW88423 standard; Protein; 928 AA.

AC AAW88423;

XX 26-APR-1999 (first entry)

DE Chlamydia pneumoniae surface exposed protein Omp10.

KW Omp10; outer membrane protein 10; surface exposed protein;

KX antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.

OS Chlamydia pneumoniae.

PN W09858953-A2.

PD 30-DEC-1998.

PF 19-JUN-1998; 98WO-DK00266.

PR 23-JUN-1997; 97DK-0000744.

PA (BIRK/) BIRKELUND S.

PI (CHR/L/) CHRISTIANSEN G.

XX Birkelund S, Christiansen G, Knudsen K, Madsen A;

PI Mygind P;

DR MPI; 1999-105610/09.

DR N-PSDB; AAX06822.

PT Species-specific test for identifying mammals infected with

PT Chlamydia pneumoniae - comprises detecting antibodies specific for

PT outer membrane proteins of C. pneumoniae or nucleic acids encoding

PT these proteins

PS Claim 7; Page 60-62; 115pp; English.

XX This polypeptide comprises the novel 98.4 kDa surface exposed

CC protein Omp10 of the human respiratory pathogen Chlamydia

CC pneumoniae. Its amino acid sequence was deduced from DNA (see

CC AAX06822) isolated from a C. pneumoniae expression library. The

CC invention provides 12 novel surface exposed proteins, Omp4-Omp15

CC (see AAW88417-28), and nucleic acid sequences encoding them (see

CC AAX06816-27). A new species specific test is claimed that is used

CC to identify mammals (including humans) infected with Chlamydia

CC pneumoniae. The test comprises detecting antibodies specific for

CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer

CC membrane proteins, especially by PCR. The proteins are also used  
 CC in the diagnosis of C. pneumoniae infection in mammals. The  
 CC nucleic acids and proteins can also be used in the immunization of  
 CC mammals, the nucleic acids being particularly useful as DNA  
 CC vaccines for eliciting in vivo expression of antigens. The  
 CC vaccines may also prevent atherosclerosis and bronchial asthma,  
 CC which are possibly associated with C. pneumoniae.

XX  
 XX Sequence 928 AA;

Query Match 30.0%; Score 1442; DB 20; Length 928;  
 Best Local Similarity 36.2%; Pred. No. 9,7e-105;  
 Matches 348; Conservative 169; Mismatches 373; Indels 70; Gaps 25;

QY 1 MRSILGFPVLFSLTL-LSY-FDTSLSATITISLTPEDSFEGDSQMERSTVQA-GGVYS 57  
 DB 1 MRSILGFPVLFSLTL-LSY-FDTSLSATITISLTPEDSFEGDSQMERSTVQA-GGVYS 60  
 QY 58 LMGDSISNVN-SALNKACFNTSGSVTFAGNHGILYFNNISGTTKEGAVLCCDPQA 116  
 DB 61 lmgdsisnvn-salnkcacfntsgsvtfagnhgilypnnisggttkegavllccdpqa 113  
 QY 117 TAR----FSGFSLFPIQSGDKEGCLYSKALMLNNYVRFEGNOSKTRGAI SGA 172  
 DB 114 aakllstfsgfslfpiqsgdkegclyskalmnnnyvrfegnosktrgaigsa 173  
 QY 173 NMTVGNVDSVFYQNAAT-FGGAIHSSGPIQTAIVNQAETFRPAONAKNGSGALSDGD 231  
 DB 174 sstlsajn-polltkahatqkgalysctgiltantlnasfseantaan-nggalyleas 231  
 QY 232 IDIDONAVYLFRENEALTAIGKGAVCCLPFGSSSTPVPVTFSDNKOQVFRNHSIMG 291  
 DB 232 sstlsajn-polltkahatqkgalysctgiltantlnasfseantaan-nggalyleas 231  
 QY 232 sstlsajn-polltkahatqkgalysctgiltantlnasfseantaan-nggalyleas 231  
 DB 232 sstlsajn-polltkahatqkgalysctgiltantlnasfseantaan-nggalyleas 231  
 QY 292 GGAITYARKLISISGCPPLFNNITSYANSONLGAIAIDGEGEISLAEKGTTFQGN--- 348  
 DB 288 ggaityarklisisgcpplfnnitstyansonlgaiaidggeislaekgttfqgn--- 347  
 QY 349 -----RTSLPFLNGFHLHLLQNAKFLKQANRNGSGTFEYDPTISE--ADGSTOININ 396  
 DB 348 kgasssgqtltnsinign-----tnakivqlrasgqntlyfydpiltunhaalsda 402  
 QY 397 GD--PKNKEVGTILFSGEK--SLANDPRFKSTIPQVNNLSAGYLVIKEGAEVYSKF 451  
 DB 403 gdlagapaygvlivfsgelkseaeaaadnlkstlqpltaggqslskysltlvaksf 452  
 QY 452 TQSPGSHVLDTGLTKLAKEDIAITGLAIDDSLSSSSTAAVYKANKQISVYDSIE 511  
 DB 463 sqspgshvlldtgltklakediaitglaidddsllsssstaaavykankqisvysie 511  
 QY 512 LISPNTAVYEDLNRMSQTPPLSLFEPGAGGSYVTV--AGDFPVSP--HFGFGGNKLAW 568  
 DB 521 lvpnsnvyedvswmpqvscitltadpanihidlaadpleknpilwgygnawalsv 560  
 QY 569 ---TGTGNKVGEEFMDKINRPPREKGNLVPNIIGWNAVNSLMQVQETHASSLQTD 625  
 DB 581 gdtatkskaactltwctktygnpuperfirtglvantlwgstfvdvrsiqglvatkyrgqetr 640  
 QY 626 GLMIDIGFNFHVSASEDNIRYRHNSGTYLVSNNETTPKHTYSMAFSQLESDOKRYANS 665  
 DB 641 glwcegisnfifhkdcklkngfthlsagyyvgaattlaadnlitaafcgfgydrthfhn 700  
 QY 686 NNEHYRYLGSYLQY--TNSLGNIFRYASRNPRVNVNGILSRRLQNPIMFHLCAVGH 743  
 DB 701 knraasyaaslhlqlatlspsellrylpyses-----eqp-vlfdaqisyls 748  
 QY 744 TTDMDKTDYANFPMVKNMNMNCAIECGSMPLVFPENGRFLQGAIPFMKLOLVYAYQGG 803  
 DB 749 kntmktyyqgpkgyessywndgatalaaslphtalsheglflhayrpfllkveasyinhds 808  
 QY 804 FKE-TTADGRFNSGSLTISVPLGIRFEKLALSDQVLYDFSSSYLPDIFRKDPSCAEAL 862  
 DB 804 fke-ttadgrfnsngsltisvplgirfeklalsdqvlydfsssyldpifrkdpseaal 862

DB 809 fkeurtlvrsfsgdlinvsvpdlftrfssrnerasyeatvlyadvyrknpdcttal 868  
 QY 863 VISGDSMLVPAHVSHAPVSGTGRYHFNDYTELLRCGSIIECRPHARNYNINGSMFR 922  
 DB 869 llnstskttgtlnstsqaglrqglfyaalspnlevsnlsmlrsgsrsvynadlqgkft 928

RESULT 9  
 AAY35052  
 ID AAY35052 standard; Protein; 930 AA.

AC AAY35052;  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX

DE Chlamydia pneumoniae surface exposed polypeptide.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN MO9927105-r2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98MO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

PA (GEST ) GENSET.

PI Griffais R;

DR WPI; 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

PS Page 940-942; Disclosure; 1912pp; English.

CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAY91990) of Chlamydia pneumoniae.  
 CC C. pneumoniae causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
 CC nucleotide sequences can also be used as immunogenic compositions,  
 CC especially where the vector directs the expression of a neutralising  
 CC epitope of C. pneumoniae.

XX  
 XX Sequence 930 AA;

Query Match 29.8%; Score 1434; DB 20; Length 930;  
 Best Local Similarity 36.6%; Pred. No. 4.2e-104;  
 Matches 344; Conservative 157; Mismatches 393; Indels 44; Gaps 21;

QY 10 LVESLTLVSVFDSLS--ATTISLTPEDSFH--GDSQNAERSYVQADVYSITGDSV 65  
 DB 8 lvsstlvlpillstltvgadalspstdfdsgsgstfcpkstadangnynylsgnvyln 67  
 QY 66 NV-DNSALNKACFNTSVSGVTFAGNHGILYFNNISGTTKEGAVLCCDPQATARFSGFS 124  
 DB 68 daqkygalgcctfctgtlgtfgykysfntvdagsnagaastadkalt--ftgs 125  
 QY 125 TLSFISLPEDIKQGG--CLYSKALMLNNYVRFEGN--NOSKTRGAI SGA NVITVGN 179  
 DB 126 nlsfiapgtlvassgkstsasagalnltngltllfqnsveannggaltaktlsisgn 185



Db 416 tgnlftgkiselseteadaadsknltskllqpvtlsggtslkhgvtlqtgafgqdsrlm 475  
OY 462 DCGTKRIASKEDEIAITGLAIDIDSLSSSTAAVTKANTANKOISVDSIEIPIPTGNAYE 521  
Db 476 dygttl-epadstlmlylnissigakak-ietkatsknltsqtlillobtftfye 533  
OY 522 DLKMRNQTFFPLSLBPGAGSVYTAGDFLPVSP-----HYGFGQNW-KLAWTGTG 572  
Db 534 nhlrnpqsydllelk--aagvtstca-----vtpgplmgekfhygqgfwplw-qtg 585  
OY 573 -NKVGEFFMDKIKYKRPKEKGLVNNILMGANVNRSLMQOVETHAASSIQTRGLMIDG 631  
Db 586 astatinfwtktylqnpereigslpnslnwalidslshymetaneglgqdratwcaq 645  
OY 632 IGFNFVHASSEDMIRYRNHNSGGVYLSVNNETPKHYTSMAFSOLFSDKXDYAVSNNEYRMA 691  
Db 646 lsnfifkdstktrgrfrhssgyvlglnhtcsdkllsaafcllqdrdrfyaakngv 705  
OY 692 YLGSYLQYTTSLGNIFFRNASRNPVNVGILSRRLQNPMLIFPLCAYGHAATNDKTDY 751  
Db 706 yggtllyqhmet-----ylslpcklirpslsyvpelr-vltsnglsyhtdnldlktky 758  
OY 752 ANPMPKNSMRNMCMAIECGGSMPLVFFENGRLFGCAIRPMKQLVYAYOGDEKETTADG 811  
Db 759 tlyptkgsagndstalefgrapicidasa-lfegymplmklyfyaqheqfkeqgtea 817  
OY 812 RFRSNGSLTISVPLGIRFEKIALSODVLYDFSFSTYIPDIFRKDPCSEALVTSQDSMLV 871  
Db 818 refgsrlnlhalpigrtdkesdcgdatcnyllygtvdlvrenpoccttlrlsagswkt 877  
OY 872 PAHNSRAHFGVSGTRHYHNDYTELLCRGSIECRPHANNYNINGSKREFE 922  
Db 878 fgnlaraqalvtragnhfcfnstfeafsgsfelrlsgrsxnvdigakxygf 928  
RESULT 11  
AAV90240  
ID AAV90240 standard; Protein: 930 AA.  
AC AAV90240;  
XX  
XX 29-AUG-2000 (first entry)  
DT  
XX Chlamydia antigen CPN100639.  
DE Chlamydia antigen CPN100639.  
XX  
XX Chlamydia antigen; diagnosis: infection; community acquired pneumonia;  
KM therapy: upper respiratory tract disease; bronchitis; sinusitis;  
XX asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma.  
OS Chlamydia pneumoniae.  
XX  
XX WO200032794-A2.  
PN  
XX  
XX 08-JUN-2000.  
PD  
XX  
XX 01-DEC-1999; 99WO-CA01147.  
PF  
XX  
XX 01-DEC-1998; 98US-0110339.  
PR 01-DEC-1998; 98US-0110340.  
XX 01-DEC-1998; 98US-0110427.  
PR 01-DEC-1998; 98US-0110428.  
XX 01-DEC-1998; 98US-0110438.  
XX  
XX (CONN-) CONNADGHT LAB LTD.  
XX  
XX  
XX Mordin AD, Oomen RP, Wang J;  
PI  
XX  
XX WPI: 2000-412339/35.  
DR N-PSDB; AAA30853, AAA30854.  
XX  
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for  
PT preventing, diagnosing and treating diseases such as community acquired  
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset

PT asthma -  
XX  
XX Claim 16; Fig 7; 174pp; English.  
CC This sequence is a Chlamydia antigen of the invention, designated  
CC CPN100639. The nucleic acids (and their complementary sequences) may be  
CC used as diagnostic agents for detecting the presence of nucleic acids  
CC encoding Chlamydia antigens in samples according to standard methods,  
CC and therefore, for diagnosing Chlamydia infections. For example, they may  
CC be used as primers and probes for diagnostic polymerase chain reaction  
CC (PCR) assays. Antisense sequences may be used to down regulate  
CC expression of the proteins and may be used to treat infections. The  
CC nucleic acids may also be used to produce the protein antigens they  
CC encode according to standard recombinant DNA methodologies. The  
CC proteins may then be used as antigens for the production of antibodies  
CC (i.e. as vaccines) for preventing infection by Chlamydia. The  
CC antibodies may also be used as diagnostic reagents for detecting  
CC infections. Chlamydia is a pathogen implicated in the development of  
CC (for example) community acquired pneumonia, upper respiratory tract  
CC disease (especially bronchitis and sinusitis, asthmatic bronchitis,  
CC adult-onset asthma and acute exacerbations of asthma in adults.  
CC  
XX  
SQ Sequence 930 AA;

Query Match 29.6%; Score 1426; DB 21; Length 930;  
Best Local Similarity 36.5%; Pred. No. 1.8e-103;  
Matches 343; Conservative 157; Mismatches 396; Indels 44; Gaps 21;

OY 10 LVFSLTLLSVFDTSL--ATTSLTPEDSFH--GDSQNAERSTVNOAGDYISLTGCVSIS 65  
Db 8 LLSLSTLVLPIIISTATYADASISIDTGDAGSGTTFKPSADANGYVLSGNVYIN 67  
OY 66 NV-DNSALNKACPNVTSVSGVTFAGNHGLEYFNNISGGTTEKCAVILCCOPQATARSQGS 124  
Db 68 dagkgtalgcctteltgdtlftgkysfintvdagsnagaastadkall--ftgts 125  
OY 125 TUSFIQSPGDIKEQG--CLYSKNALMLNNVYVRFQ--NOSKTRGAIISGANVTIVGN 179  
Db 126 nlsfiagptlvasgktslssagalnldngllfsgvnseannngaitlktltslsgn 185  
OY 180 YDSVSFYONAA-TEGCAIHSSGPIQIAVNOAERFQONTAKNNSGALYSDGDIDIDQNA 238  
Db 186 tsstlfttsnakkllggaiaysaaaslsngtlqglvnmhngely-ggallfcaasssltqns 244  
OY 239 YVIFRENEALTTAIGKGGAVCCLPRTSGSSNPVPIYFSPNKOYFERNISIMGCAIVAR 298  
Db 245 slfsgnta-tdaagkysgaiaiekyt-----eplrlsngkaltfseassvvggaltch 299  
OY 299 KLSISSGPTLFINNISYANSONLGAIAIDTGEISLSAEKGTTFQGN--RTSLP-- 353  
Db 300 gldlsaagptlffennrcgntlaagkgalaadagslisaangdltflgnltstsaeps 359  
OY 354 FLNGHLLONAKFLKQIARNGCSIEFYDPTSEADOSTQ-LNINQDPKK--EYGTILF 410  
Db 360 tmaiaylgssakltlraaaggslyfydplaasntcyascdvltngdpdsnpldygslvlf 419  
OY 411 SCGRSLANDPR--DEKSTIPQVNVNLSAGVLYIKGAETVSKFTQSPSHLVLDTGTL 467  
Db 420 sgeklisaadeakaadntslkqplalaasgtlalkgnveldivngftgtesltlmpgkxl 479  
OY 468 IASKEDIAITGLAIDIDSLSSSTAIVAKANTANKOISVDSIEIISPQNAVYEDLRMN 527  
Db 480 kadteaistklivdlsalegnkvsaleag-ankltltsplvfqdsngfnfeshltng 538  
OY 528 SQTFFPLSLBPG-AGGSVYTAGDFLPV---SPHYGFGQNMKLAWTGTGN-KVGEFFMK 582  
Db 539 artqplivftaataasdyidaillspvqtppehyggygghwaetadstasklmtwt 598  
OY 583 INKRPKEKGLVNNILMGANVNRSLMQOVETHAASSIQTRGLMIDIGNFFHVSASE 642  
Db 599 tynnperrassvpaswaftdlrltqlqmtsqaansiyqgrglwaastanflhkdksg 658







XX	Chlamydia pneumoniae.
XX	
XX	W09927105-A2.
XX	
XX	03-JUN-1999.
XX	
XX	20-NOV-1998; 98WC-IB01890.
XX	
XX	04-NOV-1998; 98US-0107078.
XX	21-NOV-1997; 97FR-0014673.
XX	
XX	(GEST ) GENSET.
XX	
XX	Griffais R;
XX	
XX	WPI; 1999-357842/30.
XX	
XX	Genome sequence of Chlamydia pneumoniae
XX	
XX	Page 942-944; Disclosure; 1912pp; English.
XX	
XX	AAV34584-Y35879 represent the proteins encoded by all the open reading
XX	frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
XX	
XX	C. pneumoniae causes respiratory disease such as pneumonia and
XX	bronchitis and is thought to be a contributing factor in heart
XX	disease, sarcoidosis, sinusitis, purulent otitis media, erythema
XX	nodosum or pharyngitis. The polypeptides encoded by the open reading
XX	frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in
XX	immunogenic compositions as vaccines. Vectors containing C. pneumoniae
XX	nucleotide sequences can also be used as immunogenic compositions,
XX	especially where the vector directs the expression of a neutralising
XX	epitope of C. pneumoniae.
XX	
XX	Sequence 927 AA;
XX	

Query Match	29.2%	Score 1405.5;	DB: 20;	Length 927;
Best Local Similarity	35.78%	Pred. No. 7.3e-102;		
Matches 343;	Conservative 166;	Mismatches 380;	Indels 71;	Gaps 24;

[illegible]

```

OY 452 TQSESHLYDLGRKRLKASKEDIAITGLAIDIDSLSSSTRAAVIKATANKAOLSTPDSIE 511
Db 463 sqspssltlmgaltletadagsslticqcftrkrdext---lkaqsaqvtlvisgs 519
OY 512 LISPGNAVEJDLRMKNSOTFPLLSLPEAGAGSVYTV--AGDFLPVSD-HYFGQGNMLTAW 568
Db 520 lvdpgnuyevdswmnpqvfascrltaddpnnihltlaadpleknplhwgygnvalsw 579
OY 569 ---IGTGKXVEEFWDKINRYKPREKEBNLYPNILMGNAVNVRSLMQVOEITHASITQTOR 625
Db 560 gedatkskaatlwtlwklygnpndperrgtlvaantlwsfydvrsizqglvaltkvrqsctr 639
OY 626 GLMIDGINFEPHVSASDENIRYRNSGCGYVLVSNNELTPKHYTSMASQULFSRDXDAVS 685
Db 640 givcegsinffhkdstklnkfgrhnsagvyvgattlasnlltaacqglfgrdrhfin 699
OY 666 NNEYRMYLGSYLYOY--TTSIGNIFRTASRNPNNVNGILSRFLQNPMLMFIHLCAYGHA 743
Db 700 knrasayaasahlqlhalisspsllylpgsae-----eqp-vlfdagisyy 747
OY 744 TNDKMTDYANPVMYKNSRRNNCMALIECGSSMPLLVFENGRLFGQAIPEMKIOLYAYOGD 803
Db 748 knmklyttqgbkesswynogcalelasslphthalshgllfhyfrfllkveasylnqds 807
OY 804 FKE-''TADGRFRSNGSLTSLISVPLGIREKTLASQDYLVDYFSEFSYIDIFRKPDSCEAL 862
Db 808 fkerntltvisfsgdglinvsvpjlfrtfrtsrnerasyaetviyadvayyrknpdcttal 867
OY 863 VISGDSMLVPAHNSRNAFVSGGCRHYFNFDYELLCLRGSIIECPHARANNINGSGFRF 922
Db 868 llnmswtvltglnstgrgigragrlfyaifsnplvtlsmelntlgsrsgvntadlvggftfqi 927

```

RESULT	15
AAV99843	
ID	AAV99843 standard; Protein; 925 AA

DT	15-SEP-2000	(first entry)
DE	Chlamydia pneumoniae processed 98 kDa outer membrane protein CPN100640.	
XX		
XX		
XX	Chlamydia; 98 kDa outer membrane protein; antigen; immunogen; infection	
KW	vacctie; antibacterial; community acquired pneumonia; bronchitis;	
KW	sinusitis; acute respiratory disease; upper respiratory tract disease;	
KW	asthma; atherosclerosis.	
XX		
OS	Chlamydia pneumoniae.	
XX		
PN	WO200032784-A1.	
XX		
XX	08-JUN-2000.	
XX		
PF	01-DEC-1999; 99WO-CA01148.	
XX		
XX	01-DEC-1998; 98US-0110439.	
PR	03-MAY-1999; 99US-0132272.	
XX		
PA	(CONN-) CONNAUGHT LAB LTD.	
XX		
PI	Murdin AD, Oomen RP, Wang J, Dunn P;	
XX		
DR	WPI: 2000-412330/35.	
DR	N-PSDB: AAA48838, AAA48839.	
XX		
PT	New polynucleotide encoding the Chlamydia 98 kilodalton outer membrane	
XX	protein, useful for preventing or treating Chlamydia infection -	
PS	Claim 16; Fig 1; 98pp; English.	
XX		
CC	The present sequence is the mature, processed form of CPN100640, the	
CC	98 kDa outer membrane protein from Chlamydia pneumoniae. Chlamydia	

CC pneumoniae is a common cause of community acquired pneumonia and upper  
CC respiratory tract symptoms and diseases, including bronchitis and  
CC sinusitis. It also has an association with atherosclerosis and asthma.  
CC The 98 kDa outer membrane protein is a C. pneumoniae-specific antigen  
CC which can confer immune protection against chlamydial infection. The  
CC nucleotide sequence encoding the protein or the protein itself may be  
CC administered as a vaccine to prevent or treat infection and they may  
CC also be used to diagnose infection. The gene encoding CPN100640 was  
CC amplified from Chlamydia pneumoniae genomic DNA by PCR.

XX Sequence 925 AA:

Query Match 29.1%; Score 1400.5; DB 21; Length 925;  
Best Local Similarity 35.9%; Pred. No. 1.8e-101;  
Matches 338; Conservative 166; Mismatches 391; Indels 47; Gaps 21;

OY 9 PLVFSLTLSVFPDLSATTSLTPEDSFHGSQNAERSYVQAGDVYSLTGDVSISSND 68  
DB 3 PLFSLSIVAA-EVLIDSSNYSY---DGSNGTTFVFTSTDAAGTYSLLSDVSFQNA 58  
OY --NSAANKACFNTSSGVTFAGNHGILFNNISSGTTKEGAVLCODPQATARESGFSTL 126  
DB 59 ALDPIASGCLFleaagdlftqgnhalktaflnags-aglvastaadknllfndfstrl 117  
OY 127 SFIOSPGDIKE--QGCLYSKNALMLNNYVFEQNSQKTKGALSGANVTIVGNYSV 183  
DB 118 SLSGCSPLISPGGALKSVGNLSITGNSQILFTQNSGDNVGNLTKNLLSGTSQFA 177  
OY 184 SFYQNA---TFCGAIHSSGPIQIAVQAEIRQAQNTAKNGSGALYSDGDDIDQNAVY 240  
DB 178 SFERNAGFTGKGGVYATGLTILNENSPGIVSFGNLAQ-SSGALYSTDCSLDNFQY 236  
OY 241 LPEENALTTAIGKGAVALCPSSGSTPVPITFSPDNKQLVFERHNSIMGGAIVARKL 300  
DB 237 IIdgnswaagaqgagalccttdkt----VLCGNKLSFTMCALYGAISGLKV 290  
OY 301 SSSGPGTLFNNISYANS-ONIGALAIPTGGEISIAEKGTTTFQGNRT---SLPELN 356  
DB 291 SLSAGPILFQNSISGSSAGGQGNINLASAGELASGALTSGLTNNGVNLTGSGTSTN 350  
OY 357 GHLLQNAKFLQANNGCSIEFYDPTS--EADGSTQLNINQDPKNK--EYTGTLIFSG 412  
DB 351 ahlldtktvtsiraatqgslfydltnpqtaastclnlnladanselegatvfy 410  
OY 413 EKSLAND---PRDFKSTIPQNVMLSAGYLIRKGAETVSKFTQSPGSHLYLDGFKLIA 469  
DB 411 EKISPTekaaavtstlirpavlaragdvlrldgvtvtfkdlcqpsrllimgdgtlisa 470  
OY 470 SKEDIAITGLAIDISLSSSSTAIVIKANTANKQISVTDSIELISPTGNAYEDLRRMSQ 529  
DB 471 keanlslnglavnlssldgtnkaa-lkteaadnlslsgtlalldtegsfyeenhkksas 529  
OY 530 TFFPLSL-EPGAGGSVTVAGDFLV---SPHYGQGNMKLAW-TGTGNKVGEFPWDKIN 584  
DB 530 TFFPLSL-EPGAGGSVTVAGDFLV---SPHYGQGNMKLAW-TGTGNKVGEFPWDKIN 584  
OY 530 TFFPLSL-EPGAGGSVTVAGDFLV---SPHYGQGNMKLAW-TGTGNKVGEFPWDKIN 584  
DB 530 TFFPLSL-EPGAGGSVTVAGDFLV---SPHYGQGNMKLAW-TGTGNKVGEFPWDKIN 584  
OY 585 YKRPPEKGLVNIIMGNVAVNRSIMOVETASSLOTRDGLWDIGIGNPFVHASSEDN 644  
DB 590 YLPSPERKSNLPLNSLWGNIFDIRSLNGLIETKSSGFEFRELWLSGLANFIYRDSMPLR 649  
OY 645 IRYRHNSGCVLSVNNELTPKHYTSMASFQSLFRDKRDYAVSNNEFRMYLGSYLYQYTTSL 704  
DB 650 hgftrhsaggylatitpaedqitafacqlfardrnhltgknhgdtlygaslyfhnlegl 709  
OY 705 GNTFRY----ASRNPVNVNGILSRRLQNPFLMFHLCATGHAINDMKTDYANFPWKNS 760  
DB 710 fdianflwkatrap---wvlsesqilpils-fdakfsylhcdnmktytldnsilkgs 764  
OY 761 WRNRCWAIEGGSMPLLVFNENGRLPQGAIRPEMLQLYVAYOGDFKETTADGRFNSGLT 820  
DB 765 wrndatcaadlgaslp-fvisvpyllkevepfvkvqylyanqgdlyernaegrafnksell 823

OY 821 SISVPIGIREKIALSDVLYDFSFYSYIPDIFRKDPSCEALVISGDSMLVPAHVSRAH 880  
DB 824 nveilpylterdskekydltlmylidayrtnpkcqtslasdanmmygtlnlarqg 883  
OY 881 FVSGGTRHYFNDYTELLCRGSIECRPHANRYNINGSKRFR 922  
DB 884 fsyraanhfqnphmeilfggfatevrssrnytlngsktcf 925

Search completed: October 2, 2001, 03:26:29  
Job time: 30221 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2001, 03:27:22 ; Search time 52.09 Seconds  
(without alignments)  
364,452 Million cell updates/sec

Title: US-09-446-677b-6

Perfect score: 4814

Sequence: 1 MRFSLGPFPLVFLSLTLVSF.....IECRPHARNYNINGSKFRF 922

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/6C.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/6C.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192.5	4.0	1338	2	US-08-728-470-9
2	192.5	4.0	1338	4	US-08-719-641-9
3	192.5	4.0	1599	2	US-08-617-697-9
4	192.5	4.0	1536	1	US-08-038-682-2
5	192.5	4.0	1536	1	US-08-302-832-2
6	192.5	4.0	1536	2	US-08-530-198-2
7	192.5	4.0	1536	2	US-08-469-880-2
8	192.5	4.0	1536	2	US-08-728-470-2
9	192.5	4.0	1536	2	US-08-617-697-2
10	192.5	4.0	1536	4	US-08-719-641-2
11	178.5	3.7	1529	2	US-08-728-470-10
12	178.5	3.7	1529	4	US-08-719-641-10
13	178.5	3.5	1477	1	US-08-038-682-4
14	170.5	3.5	1477	1	US-08-302-832-4
15	170.5	3.5	1477	2	US-08-530-198-4
16	170.5	3.5	1477	2	US-08-469-880-4
17	170.5	3.5	1477	2	US-08-728-470-4
18	170.5	3.5	1477	2	US-08-617-697-4
19	170.5	3.5	1477	4	US-08-719-641-4
20	170.5	3.5	1612	1	US-08-169-927-2
21	167.5	3.5	1600	2	US-08-617-697-10
22	156.5	3.3	1838	4	US-09-120-663-2
23	150.5	3.1	1160	3	US-08-808-599A-24
24	148	3.1	1394	4	US-08-296-791-2
25	148	3.1	1394	5	PCT-US95-10661A-2
26	147	3.1	1026	2	US-08-614-377A-7
27	147	3.1	1026	4	US-09-142-648B-7

28	145	3.0	1026	1	US-08-194-290-7	Sequence 7, Appl
29	145	3.0	1545	5	US-08-296-791-4	Sequence 4, Appl
30	145	3.0	1545	5	PCT-US95-10661A-4	Sequence 4, Appl
31	144.5	3.0	2353	4	US-09-377-155-33	Sequence 33, Appl
32	144.5	3.0	2353	4	US-08-913-942-4	Sequence 4, Appl
33	141.5	2.9	1912	1	US-08-409-995-4	Sequence 4, Appl
34	141.5	2.9	1912	3	US-08-685-467-4	Sequence 4, Appl
35	140.5	2.9	1222	2	US-08-682-517-15	Sequence 15, Appl
36	140.5	2.9	1252	2	US-08-682-517-9	Sequence 9, Appl
37	138.5	2.9	599	4	US-09-377-155-15	Sequence 15, Appl
38	138.5	2.9	1541	4	US-08-296-791-3	Sequence 3, Appl
39	138.5	2.9	1541	5	PCT-US95-10661A-3	Sequence 3, Appl
40	135	2.8	2123	4	US-08-968-685A-10	Sequence 10, Appl
41	133.5	2.8	592	4	US-09-377-155-17	Sequence 17, Appl
42	131.5	2.7	598	4	US-09-377-155-5	Sequence 5, Appl
43	131.5	2.7	910	4	US-08-460-269C-2	Sequence 2, Appl
44	130.5	2.7	591	4	US-09-377-155-11	Sequence 11, Appl
45	130.5	2.7	591	4	US-09-377-155-21	Sequence 21, Appl

#### ALIGNMENTS

RESULT 1  
US-08-728-470-9  
Sequence 9, Application US/08728470  
Patent No. 5928651  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
SUITE: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,470  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
TELEPHONE: (703) 415-0810  
TELECOMMUNICATION INFORMATION:  
1038-633  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1338 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-728-470-9



```

Db 706 IS-----NKGSDNLIDKSDAEIO-IGNIS-0KEGMLTSSDKVNTNQ 749
QY 347 -----GNFTSLPFLNGIHLONAKFLK-----QARNCSIEFYDPT 384
Db 750 ITIKAGVEGGRSDSEANANLTIOTKELKLAGDLNISGFNKAETFAKNSDILIGNASG 809
QY 385 SEADG-----ST---QLNINGDPK-----NKEYTITLFSGKSLAND 419
Db 810 GNAADAKKVTEDKYDKSKISTDGHNVTLNSEVKTSNGSSNAGNDNSTLITISAKDVTYNN 869
QY 420 PRDEKSTIPONVNLISAGLYIKEGAETVSKFTQSPGSHLVLDGTLKLIASKEDIAITGL 479
Db 870 VTSKRTI---NISAAGNVTTKEGTTIN-----ATTG- 898
QY 480 AIDIDSSSSSTAIVIKANRANKOISVDSIELISPTG---NAYEDLRMNSOTFPLLSL 536
Db 899 -----SVEVTAONGTIGKNTTSQNVVTATENAVITENAVINATSGVTNISTKGDIKG 953
QY 537 EPAGGSVTVTA-GDFLPVSPHYG-----FOGNNKLATGTGN---KGEF 578
Db 954 IESTSGNVNITTAGSNTLKVSNTIGQDVTYADAGALTTAGSTISATYGANITTKGD- 1012
QY 579 FMDKINKPRPEKEGNL-----VPIILGNANVNSLMQVOTHTASSLOTDEGLW 628
Db 1013 -----INKVE-SSSGSVTLVATGATLAVGNIS-GNTVYTTA-----DSGKLTSTVGST 1059
QY 629 IDGI-----GNFHVASSEDNIRRH-----NSGGYLVSVNNEITP 664
Db 1060 INGNNSVTTSSQSDIEGTIGTNTVNTASTGDLTIGNSAKVEAKNGAATLTAESEGLTT 1119
QY 665 KHTSMAFS--QLFSRDKDYAVSNN 687
Db 1120 QTGSSTSSNGQTTTAKDSSIAGN 1144

```

```

RESULT 3
US-08-617-697-9
; Sequence 9, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Maltare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Betkieseser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1599 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-617-697-9

```

```

Query Match 4.0%; Score 192.5; DB 2; Length 1599;
Best Local Similarity 20.0%; Pred. No. 2,4e-07;
Matches 173; Conservative 125; Mismatches 278; Indels 289; Gaps 42;

```

```

QY 13 STLLSVDFTSLAFTTSLPDESFHGDSONAERSY-----NQAGDVYSLTGD----- 61
Db 639 TLNISTGVDSIMKAPKVS-----FYRDK---GRTYNNVTLNVTSGSKFNLISDTSG 690
QY 62 ---VSTNVNNSAL--NKAFENVTSGS-----VTFAGNHGILYFN---MISSGT 103
Db 691 STGPSIRMAELNGITTFENKAFENIAQGSTANFSIKASIMPKSNANVALFNEDISVSG-- 748
QY 104 KEGAVLCCODPQATARPSGFTLSFISPG-----DIKROGCLYSKNAL 147
Db 749 -----GSVNFKLASSNSNIOTPGVYIILKSQFNVSGSSTLWLKAEGS--TETAF 794
QY 148 MLN-----NYVR-PEONOSKTKGAGISGANTYVGN-----YDSVS 184
Db 795 SIENDLNLNATGNTITRQEGDHSRVNKGVAKKNTTFKGNITTFGSKATTEIKGNVT 854
QY 185 FYQNA-ATFGALHSSGPTLOAVNQAEIRAOATKAG---SGGALYSQDDIDONAY 239
Db 855 INKNTATLNGA-----NFAENKSPNLNAGVINGNLTGASTIINIGNLTFSKGAN 907
QY 240 V-----LRENEALTTAIGKGAVCCLPTSGSSTPPIYFESDNK-OLVERN 286
Db 908 LOAITNTFENVAGSFEDNGNSNISIAKGAK--FKDINNTSSLINTTNSOTTYRTIIRGN 965
QY 287 HSIIMGGAITARKLISSSGPTLEFINNIYSANSONLGAIAIDTGEISLSAEKGTITFO 346
Db 966 IS-----NKGSDNLIDKSDAEIO-IGNIS-0KEGMLTSSDKVNTNQ 1009
QY 347 -----GNFTSLPFLNGIHLONAKFLK-----QARNCSIEFYDPT 384
Db 1010 ITIKAGVEGGRSDSEANANLTIOTKELKLAGDLNISGFNKAETFAKNSDILIGNASG 1069
QY 385 SEADG-----ST---QLNINGDPK-----NKEYTITLFSGKSLAND 419
Db 1070 GNAADAKKVTEDKYDKSKISTDGHNVTLNSEVKTSNGSSNAGNDNSTLITISAKDVTYNN 1129
QY 420 PRDEKSTIPONVNLISAGLYIKEGAETVSKFTQSPGSHLVLDGTLKLIASKEDIAITGL 479
Db 1130 VTSKRTI---NISAAGNVTTKEGTTIN-----ATTG- 1158
QY 480 AIDIDSSSSSTAIVIKANRANKOISVDSIELISPTG---NAYEDLRMNSOTFPLLSL 536
Db 1159 -----SVEVTAONGTIGKNTTSQNVVTATENAVITENAVINATSGVTNISTKGDIKG 1213
QY 537 EPAGGSVTVTA-GDFLPVSPHYG-----FOGNNKLATGTGN---KGEF 578
Db 1214 IESTSGNVNITTAGSNTLKVSNTIGQDVTYADAGALTTAGSTISATYGANITTKGD- 1272
QY 579 FMDKINKPRPEKEGNL-----VPIILGNANVNSLMQVOTHTASSLOTDEGLW 628
Db 1273 -----INKVE-SSSGSVTLVATGATLAVGNIS-GNTVYTTA-----DSGKLTSTVGST 1319
QY 629 IDGI-----GNFHVASSEDNIRRH-----NSGGYLVSVNNEITP 664
Db 1320 INGNNSVTTSSQSDIEGTIGTNTVNTASTGDLTIGNSAKVEAKNGAATLTAESEGLTT 1379
QY 665 KHTSMAFS--QLFSRDKDYAVSNN 687
; : : : : :

```

DB 1380 QTGSSITSSNGQTTLAKDSSIAGN 1404

RESULT 4

US-08-038-682-2

Sequence 2, Application US/08038682

Patent No. 5549897

GENERAL INFORMATION:

APPLICANT: BARENKAMP, STEPHEN J

APPLICANT: ST. GEME III, JOSEPH W

TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/038,682

FILING DATE: 16-MAR-1993

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: BERKSTRESSER, JERRY W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-293

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0810

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1536 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-038-682-2

Query Match 4.0%; Score 192; DB 1; Length 1536;

Best Local Similarity 20.7%; Pred. No. 2.5e-07;

Matches 174; Conservative 119; Mismatches 276; Indels 270; Gaps 39;

QY 12 FSLTLLSVFDTLSATTITSLTPEDSPFSDSONABRSYVQAG----- 53

DB 676 FNTIT-----DSKSDSAGITLTPYNNLNGISFNKOTFFNVRNARVNDIAPIGINKYSS 731

QY 54 -DVSYLTVGVVIS--NVDSAL-----NKACFNTSGSVTEAGNHGLEYFN 96

DB 732 LNVASFGNISVSGGSDVDEFTLLASSNVQTPGVINISKYFNVTSG-----SLRFK 783

QY 97 NISSGTTKEBAVLCDDPQATAFSGFS-FLSTFQSGDKEGGLYSKAMALNNYV 155

DB 784 --TSGSTKGTGFST--EKDITLNTAGNTLLQVEGTGIGIGIVAKKNITEGGINTF 838

QY 156 RFQONOSKTKGAI--SGANVTIVG--NYDSVSFYQNAATF--GGAISHSGPLQIAVNOAEI 211

DB 839 GSKKATTEIGENVTINNANVTILGSPEDN--HQRPLTIKKIVITINSGLTLAGNIVNI 895

QY 212 RFAQNTAKNGSGALYSDGIDIDQANAYVLFRENEALTT-AIGKGAVCCLPTSGSSTPV 270

DB 896 A-----GNLTVEESNA-----NFKAITNFTFNVG----- 919

QY 271 PLVTSDNKQVLEVRHNSIMGCAIT-----AKRLST-----SSGGPLFTINNISTYANSQ- 320

DB 920 ----LEDNKG--NSNISIAGKAREFDKIDNSKNLSITTNSSSTYRTIISGNTNNKNDL 972

QY 321 ---NLGGAIAIDTGGETS-----LSAEKGTTFQ-----GNRTSLPLNGIHLQ 362

DB 973 NITNEGSDTEMOIGDVSQKSEGNTITSSDKINITKOTIKIAGVDGEUSDADANNANLTI 1032

QY 363 NAFELKLOARNNGSIEFYDITSEADGSTOLNINQDPKNEYT---GTILFSEKSLAND 419

DB 1033 KTKELKL-----TDLNITSGFNR-AEITAKDGSDLTIGNTNSADG 1071

QY 420 PROFKSTIPO--NVNLSAGLYIKEGAEVTVSKFTQSPGSHVLVDLGTIKLIAKEDAIAT 477

DB 1072 TNARKVTFNQVKDSKISA-----DGKRYLHSHKRVETSGSN-----NTESSDN 1115

QY 478 GLAIDID-----SLSSSTAIVIKANTANKOISVYDSIELSPGNAYEDL 523

DB 1116 MAGLTIDAKVYNNNTITSKAVSISATSEITTKTGTTINATGNVEITAOQS----- 1170

QY 524 RMRSQTFPILLSPEGAGSVTVTA-----GDFLPVSPHYG-----FOGN 563

DB 1171 -----ILGGIESSGGSVTLTALGALAVSNISGNVTVTANSGALTYTLAGSTIKGT 1221

QY 564 WKLAWTGTGNKV-----EFWDKINYRPRPK-EGNL-----VPNIMGN 603

DB 1222 ESYTTSQSGDIGTISGTVKAEISLTQSNKIKATGTGAATGATGTGTTSG 1281

QY 604 AVVNRSLMQVQETHASSLQTRGLMIDIGNPFVSASEDNIRYRHSNGSVLNVNNEIT 663

DB 1282 TVAVTA-----NAGLT-----YNGAEINNTBEAATLTSSGULTTEASHSHIT 1325

QY 664 PKHTSMASFQSFSDKDAVSNNEYRMVGLSYLYOTTSLSGNIFFRYSANPNVNVGIL 722

DB 1326 ----SAGQVNLASQDSVAGSINAAVTL-----NTGTGLTVVKGSNINATSGTL 1372

RESULT 5

US-08-302-832-2

Sequence 2, Application US/08302832

Patent No. 5603938

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: Of No. 5603938-Typeable Haemophilus

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/302,832

FILING DATE: 16-SEP-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US pct/us93/02166

FILING DATE: 16-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-404

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1536 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-302-832-2

Query Match 4.0%; Score 192; DB 1; Length 1536;  
 Best Local Similarity 20.7%; Pred. No. 2.5e-07;  
 Matches 174; Conservative 119; Mismatches 276; Indels 270; Gaps 39;

```

QY 12 FSLTLLSVPTSLSATITSLTPEDSFHDSOMNERSYNVQAG----- 53
DB 676 FNLT-----DSRGSDSAGTLQPYNLNGISFNKDTFVNERNAVNFDIKAPIGINKYSS 731
QY 54 -DYSLGDSIS--NYDSAL-----NKACFNVTSGSVTFAGNHGLYFN 96
DB 732 LNAASFNGNISVSGGSDVDTLLASSSNVQTPGVVINSKYFNSTGS-----SLRFK 783
QY 97 NISSGTRKEGAVLCCDPQATARESGFS-TLSFTQSPGDIKEOGLSKNALMLNNYV 155
DB 784 --TSGSTKTGFSI--EKDLTLANATGNITLLQVEGDMIGKIVAKKNTTFEGGNITF 838
QY 156 RFEONSKRTGAGAI--SGANVTYV--NYDSVSYQNAATF--GGAIHSSGPIQIAVNOAEI 211
DB 839 GSKRAVTEIGENVNTINNANVTLLGSDFDN--HQKPLTIKKDVIINSGLNLTAGNIVNI 895
QY 212 RFAONTAKNSGGALYSDGIDIDQNAVYLFRENEALTT-AIGKGAVALCLPTSGSSTPV 270
DB 896 A-----GNLTVESNA-----NEKAITNFTFNVG----- 919
QY 271 PIYTFSDNKOVLVERHNSINGGAIY-----ARKLSI-----SSGGPILFINNISYANSQ- 320
DB 920 ----LFDNKG--NSNISIAGKARFMDINSKMLSTTTSSSTYRITLISGNITNKNGDL 972
QY 321 ---NLGGAIAIDTGEIS-----LSAEKGTITFQ-----GNRTSLPFLNGIHLQ 362
DB 973 NITNEGSDTEMOJIGDVSQKEGNLTISDKINITKQITIRAGVDGENSDSDATNANMLTI 1032
QY 363 NAFELKIQANNGSIEHYDITTSADOSTOLNINDGPKNEYT---CTIFSGSKSLAND 419
DB 1033 KTKELK-----TQDLNLSGFENK-AETIAKDSDLTIGTNTSADG 1071
QY 420 PRDFKSTIPQ--NVNLSAGYLVKEGAELTVSKFTQSPGSHLYDLCTKLASKEDIAIT 477
DB 1072 TNMKKVTFFNQVKSKISA-----DGHKVTILHKSVEITSGSN-----NNTEDSSDN 1115
QY 478 GLAIDID-----SLSSSTAIVIKANTANKOISVTDISTELISPTGNAYEDL 523
DB 1116 NAGLTIDAKNVNANNNTSHKAVSISATSEITTKTGTINATGANEITAGTGS----- 1170
QY 524 RMNRNSQFFPLLSLEPGAGSVTVTA-----GDELVPSPHYG-----FGN 563
DB 1171 -----ILGIESSSGSVTLTATEGALAVSNISGNITVTVANSALTTLAGSTIGT 1221
QY 564 WKLTAMTGTGNKVG-----EFFWDKINYPKRPK--EGNL-----VPNILMGN 603
DB 1222 ESVTTSQSDIDIGTISGTVVEVKATESLTTQNSKIKATGAEANVTASATGICGTISGN 1281
QY 604 AVNVRSLMOVQETHASSLQTRGLMIDICGNFHVASASEDNIRRHNSGCVLVSNNET 663
DB 1282 TVNVTAT-----NAGDLT-----YNGCAEINATECAALTTTSSGLTTEASSHIT 1325
QY 664 PKHTSMAFSGQLSRDKDYAVSNNEVRYGLSVLYQTTSLGNIERFASNPNNVNL 722
DB 1326 ----SAGQVNLSDQDSVAGSINAAVTL-----NTTGILLTVKGSINATNSGTL 1372

```

RESULT 6

US-08-530-198-2

Sequence 2, Application US/08530198  
 Patent No. 5869065  
 GENERAL INFORMATION:  
 APPLICANT: BARENKAMP, STEPHEN J  
 APPLICANT: ST. GEME III, JOSEPH W  
 TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
 TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Shoemaker and Mattare, Ltd  
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 STREET: Bldg. 1  
 City: Arlington  
 State: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202-0286  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/530,198  
 FILING DATE: 13-DEC-1995  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BERKSTRESSER, JERRY W  
 REGISTRATION NUMBER: 22,651  
 REFERENCE/DOCKET NUMBER: JWB-1186  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 415-0813  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1536 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-530-198-2

Query Match 4.0%; Score 192; DB 2; Length 1536;  
 Best Local Similarity 20.7%; Pred. No. 2.5e-07;  
 Matches 174; Conservative 119; Mismatches 276; Indels 270; Gaps 39;

```

QY 12 FSLTLLSVPTSLSATITSLTPEDSFHDSOMNERSYNVQAG----- 53
DB 676 FNLT-----DSRGSDSAGTLQPYNLNGISFNKDTFVNERNAVNFDIKAPIGINKYSS 731
QY 54 -DYSLGDSIS--NYDSAL-----NKACFNVTSGSVTFAGNHGLYFN 96
DB 732 LNAASFNGNISVSGGSDVDTLLASSSNVQTPGVVINSKYFNSTGS-----SLRFK 783
QY 97 NISSGTRKEGAVLCCDPQATARESGFS-TLSFTQSPGDIKEOGLSKNALMLNNYV 155
DB 784 --TSGSTKTGFSI--EKDLTLANATGNITLLQVEGDMIGKIVAKKNTTFEGGNITF 838
QY 156 RFEONSKRTGAGAI--SGANVTYV--NYDSVSYQNAATF--GGAIHSSGPIQIAVNOAEI 211
DB 839 GSKRAVTEIGENVNTINNANVTLLGSDFDN--HQKPLTIKKDVIINSGLNLTAGNIVNI 895
QY 212 RFAONTAKNSGGALYSDGIDIDQNAVYLFRENEALTT-AIGKGAVALCLPTSGSSTPV 270
DB 896 A-----GNLTVESNA-----NEKAITNFTFNVG----- 919
QY 271 PIYTFSDNKOVLVERHNSINGGAIY-----ARKLSI-----SSGGPILFINNISYANSQ- 320
DB 920 ----LFDNKG--NSNISIAGKARFMDINSKMLSTTTSSSTYRITLISGNITNKNGDL 972
QY 321 ---NLGGAIAIDTGEIS-----LSAEKGTITFQ-----GNRTSLPFLNGIHLQ 362

```



```

Db 973 NITNESDPEMOIGDVSQKEGNLTISDKINITKOTIKAGVGENSDATNANLTI 1032
QY 363 NAKFLKQARNCGSIEFYDPTSEADGSTOLNINGPKNKEY---GTLIFSGEKSAND 419
Db 1033 KTKELK-----TODLNISGFENK-AETIAKDGSDLTIGNTNSADG 1071
QY 420 PRDFKSTIPD--NVNLSAGYLVKEGAEVTVSKFTQSPGSHVLVDLGTKLIAKEDIAT 477
Db 1072 TNKKKVTFFNQVKSKISA-----DGHKVTLHSHKVTESGSN-----NNTEDSSON 1115
QY 478 GLAIDID-----SLSSSTAAYIKANTANKOISTVDSIELSPGNAYEDL 523
Db 1116 NAGLTIDAKNVTYNNNTSHKAVSISATSGEITTKGTITNATGAVETIAQGS----- 1170
QY 524 RMNSQTFPLSLPEPAGGSVTYTA-----GDFLPVSPHYG-----FOGN 563
Db 1171 -----ILGIESSSGSVTLTATEGALAVNISGNVTYVANSALTTLAGSTIKGT 1221
QY 564 WKLAWTGTGNKVG-----EFFWDKINYPREP-KGNL-----VPNLKGN 603
Db 1222 ESYTSSQSGDIGTISGTVKATESLITQSNKIKATGEGANTYATGTTGGTISGN 1281
QY 604 AVNVRSLMQVQETHASSLQDRGLMIDIGNPFHVASSEDNIRHNSGGYLVSNNEIT 663
Db 1282 TVAVTA-----NAGDLT-----VONGAEINATEGAALITSSGKLTTEASHIT 1325
QY 664 PKHYTMAFSQLEPSRDKDVAVSNNEYRMVLYGYTTSIGNIFRASNPNVNVGIL 722
Db 1326 ----SAKGOVNLSDQGSVAGSINAAVTL-----NTGTLLTVKGSINATSGTL 1372

RESULT 7
US-08-469-880-2
: Sequence 2, Application US/08469880
: Patent No. 5876733
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J.
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OR INVENTION: of No. 5876733-Typeable Haemophilus
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/469,880
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstreser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-516 MIS.vg
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813

```

```

: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1536 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-469-880-2

Query Match 4.0%; Score 192; DB 2; Length 1536;
Best Local Similarity 20.7%; Pred. No. 2.5e-07;
Matches 174; Conservative 119; Mismatches 276; Indels 270; Gaps 39;

QY 12 ESLTLVFPDLSLATTISLTFPDSFHGDSQNAERGVNOAG----- 53
Db 676 FNLIT-----DSKSDSAGITLQPYNLNGISFNKDTFVNERNAKRVFDKAPIGIKYSS 731
QY 54 -DYVSLTGDVVIS--NVNLSAL-----NKAQFNYSQSVTFAGNHGLYFN 96
Db 732 LNVASFNGNISVSGGGSYDPTLLASSNVQTPGVVINSKRYFNSTGS-----SLRFK 783
QY 97 NISSGTTKAGAVLCCDDPQATNAPSGFS-TLSFTQSPGDIKEGGLYSKNAALMLNNYV 155
Db 784 --TSGSKTKGFSI--EKDLTINATGNTLLQVEGTDGIGIYAKKNITFEQGNITF 838
QY 156 RFQONOSKTKGAI--SGANVTIVG--NYDSVSYQNAATF--GGAHSSGPIQAVNQAEI 211
Db 839 GSKRAYTEIGANTYNNNNANVTILIGSDPDN--HOKPLTIKRYIINSGTLTGAGIYVN 895
QY 212 RFAONTAKNNGSGGALYSDGIDIDONAYVLEFNEALTT-AIGKGAVCCLPYSGSSTPY 270
Db 896 A-----GNLTVESNA-----NFKAITNFTFVNG----- 919
QY 271 PIYTFSDNKQLYVERNHSTMGGAIF-----ARKLSI-----SSGGPTLFINNISVANSQ 320
Db 920 ---LFDNKG--NSNISTAKGARRKIDNSKNLSITNTSSSTYRITLISGNTTKNGDL 972
QY 321 --NLGALAIIDTGEIS-----LSAEKGTTFP-----GNRTSLPFLNGIHLQ 362
Db 973 NITNESDPEMOIGDVSQKEGNLTISDKINITKOTIKAGVGENSDATNANLTI 1032
QY 363 NAKFLKQARNCGSIEFYDPTSEADGSTOLNINGPKNKEY---GTLIFSGEKSAND 419
Db 1033 KTKELK-----TODLNISGFENK-AETIAKDGSDLTIGNTNSADG 1071
QY 420 PRDFKSTIPD--NVNLSAGYLVKEGAEVTVSKFTQSPGSHVLVDLGTKLIAKEDIAT 477
Db 1072 TNKKKVTFFNQVKSKISA-----DGHKVTLHSHKVTESGSN-----NNTEDSSON 1115
QY 478 GLAIDID-----SLSSSTAAYIKANTANKOISTVDSIELSPGNAYEDL 523
Db 1116 NAGLTIDAKNVTYNNNTSHKAVSISATSGEITTKGTITNATGAVETIAQGS----- 1170
QY 524 RMNSQTFPLSLPEPAGGSVTYTA-----GDFLPVSPHYG-----FOGN 563
Db 1171 -----ILGIESSSGSVTLTATEGALAVNISGNVTYVANSALTTLAGSTIKGT 1221
QY 564 WKLAWTGTGNKVG-----EFFWDKINYPREP-KGNL-----VPNLKGN 603
Db 1222 ESYTSSQSGDIGTISGTVKATESLITQSNKIKATGEGANTYATGTTGGTISGN 1281
QY 604 AVNVRSLMQVQETHASSLQDRGLMIDIGNPFHVASSEDNIRHNSGGYLVSNNEIT 663
Db 1282 TVAVTA-----NAGDLT-----VONGAEINATEGAALITSSGKLTTEASHIT 1325
QY 664 PKHYTMAFSQLEPSRDKDVAVSNNEYRMVLYGYTTSIGNIFRASNPNVNVGIL 722
Db 1326 ----SAKGOVNLSDQGSVAGSINAAVTL-----NTGTLLTVKGSINATSGTL 1372

RESULT 8
US-08-728-470-2

```

Sequence 2, Application US/08728470  
Patent No. 5928651  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,470  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:  
US-08-728-470-2

Query Match 4.0%; Score 192; DB 2; Length 1536;  
Best Local Similarity 20.7%; Pred. No. 2.5e-07;  
Matches 174; Conservative 119; Mismatches 276; Indels 270; Gaps 39;

QY 12 FSLTSLVPTSLSATITSLTPEDSFHSDSONAKRSYVQAG----- 53  
DB 676 FNLTLT---DSKSDSGAGTLTOPYNLNGISFNKDTFVVERNAVNFIDIAPIGINKYSS 731  
QY 54 -DVSLSLGDVSI--NVDSNAL-----NKACFNYSVSATYAGNHGILYFN 96  
DB 732 LNVASFNGNISVGGGVDFTLLASSNVQTPGVINSKIFNVSTGS-----SLRFK 783  
QY 97 NISSGTTKEGAVLCCDDPQATAPSGFS-TLSFTQSPGDIKEOGCLYSKNAMLNNYV 155  
DB 784 --TSGSTKTGTFSI--EKDITLNTAGNITLLQVEGTGDMIGKIYAKKNITFEQGNITF 838  
QY 156 RPEONOSKTKGAI--SGAVTIYG--NYDSVSRYQNAATF--GGAHSSGGLQIAVNAEI 211  
DB 839 GSKRAVTEIGSNVTINNANAVTLIGSDFDN--HQKPLTKKDVITINSGLTAGGNIVNI 895  
QY 212 RFAONTAKNGSGALYSDGDDIDONAYVLFRENEALTT-AIGKGAVVCLPTSGSSTPV 270  
DB 896 A-----GMLTVESNA-----NFKATLNTFTFNVG----- 919

QY 271 PIYTFSDNKOLVFERNNHSMCGGAIY-----ARKLST-----SSGGPLTFINNISYANSQ- 320  
DB 920 ---LFDNKG---NSNLSIKAGAREFDIDNSKNLSTTSSSYRTILSGNITNNKNGDL 972  
QY 321 ---NLGGAIAIDTGEIS-----LSAEKOTIYFO-----GNRTSLPFLNGIHLQ 362  
DB 973 NITNEGSDTEKQIGDVSQKRGNTLISDKINIKQITIRAGVGEKNSDSATNANANIT 1032  
QY 363 NAKFLKLOARNGCSIEFYDPTTSEADSGTOLINGDPKNEYT---GTLFSGEKSLAND 419  
DB 1033 KTKELKL-----TODLNISGFNK-AEITAKDSDLTIGMTNSADG 1071  
QY 420 PROPKSTIYPO--NWNLSAGYIVKEGAIVYSKTSQSPGSHLVLDLTKLIASGEDIAT 477  
DB 1072 TNAKVTYFNQVKSRIKA-----DGHKVTLHKSVEITSGN-----NNTEDSSDN 1115  
QY 478 GLAIDID-----SLSSSTAAYIKANTANKQISVTDIELTSPGNAYEDL 523  
DB 1116 NAGLITDAKNVTYNNNTSHKAVSISATSGEITTKGTITNATGNVEITRQTS----- 1170  
QY 524 RMRSQTFPLISLEPGAGSVYVTA-----GDPLVSPHYG-----FGGN 563  
DB 1171 -----ILGIESSSGSVTLTATGALAVSNISGNTVYVANSALITTLAGSTIKGT 1221  
QY 564 WKLAWTGTGNKV-----EPFMDKINVKRREP--EENL-----VNIILMGN 603  
DB 1222 ESVTSSQSDIGTIGTIGGVEVAKATESLTQOSNKRKATTEGANVYSAMGTIGCTISGN 1281  
QY 604 AVNYSRLMOVQETFASSLQTDRCGLMIDIGNEFHVASSEDNIRRHNSGCVLVSNNET 663  
DB 1282 TVAVTA-----NAGLIT-----VNGAELNTEGATLTITSSGKLTITASSHIT 1325  
QY 664 PKHYTNAFSQLEFSRDKDVAVSNNERYMVLGSLYQYTTSLGNIERYASRNPVNVGIL 722  
DB 1326 ---SAGOVNLSAODGSVAGSINAAMVTL-----NTGTTLTVKGSNINATSGTL 1372

RESULT 9  
US-08-617-697-2  
Sequence 2, Application US/08617697  
Patent No. 5977336  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/617,697  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 05-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651

```

: REFERENCE/DOCKET NUMBER: 1038-557
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1536 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: OS-08-617-697-2

```

Query Match	4.08;	Score 192;	DB 2;	length 1536;
Best Local Similarly	20.78;	Pred. No. 2.5e-07;		
Matches 174;	Conservative 119;	Mismatches 276;	Indels 270;	Gaps 39;

OY	12	ESLTLSEVDPNLSATTLTSLPDESFHDDSOAKESYVNOAG-----	53
Db	676	FNLT-----DSRGSOSACITLQPYLNGLISFKDITTFENRARNRPNEDIKAPIGINKYSS	731
OY	54	-DYVSLNGDVSAIS---NYDNSAL-----NKACFNVTSGSYTFAGNHHGLYFN	96
Db	732	LNYSAFNGNISVSGGSGVDFTLLASSNVQPPGVYINSKYFNVSIGS-----SLRFX	783
OY	97	NISSTTEGAVLCCODPQARFSGFS-TLSFIOSPGDIKGCGCLYSKNAIMLNNYV	155
Db	784	-TSGSYTFGFSI---EKDLTLNAGGNTLTLQVBTGDMICKGIVAKNITFECCNTFE	838
OY	156	RFEQOSKTKGAI--SGANVTIVG-NDVSIFYONATPF-GGAHSSGPIQIAVNOAET	211
Db	839	GSRAVTEIEGVNTINNANANTLISDDPN---HCKPLTIKDDYIINSNGLNAGNIYAI	895
OY	212	RFAONTAANGSGGALYSPODIDONAYVLFRENEALTT-AIGKGAVCCLEPTSSGSPV	270
Db	896	A-----GNLTVEBNA-----NFKALITFETNVG-----	919
OY	271	PIVPSDNKOLVFERNHSMGGALY-----ARKSI-----SSGGPTLEINNYSVANSQ-	320
Db	920	-----LEFDKKG---NSNISIAGGAFKIDINSKNLSTITNSSYTRTIISGNTTKNGDL	972
OY	321	--NLGAIIDAIDTGEIS-----LSAEKGTITPQ-----GNRTSLPLINGIHLQ	362
Db	973	NITNGSDTEMQIGGDVSOKECNGLTISDDKINITQITIKAGVDGENSDSATNNANLTI	1032
OY	363	NAKFLKLOARNGCSIEFYDPTISEADGSTOLINDDPNKKEYT---GTLISEGKSLAND	419
Db	1033	KTKELK-----TODLINSGFNK-AEITARDGSDLTIGNTNSADG	1071
OY	420	PRDEFKSTIPQ--NVNLSAGYLVEKGAETVSKFPOSGSHVLVDLGRKLIAKSEDIAT	477
Db	1072	TNAKVTFPNQVKDSISA-----DGHVTLIASKVETSGSN-----NNTEDSDSN	1115
OY	478	GLAIDID-----SLSSSTPAAYIKANTANKQISVTDSEIELISPTGNAEYDL	523
Db	1116	NAGLTIDAKANTVNNNITSHKAVSISATSGETITTGTTINATGNVEITTAOTGS----	1170
OY	524	RMRSQTFPLLSLEGAGAGSYTVTA-----GDFLPVSPHYG-----POGN	563
Db	1171	-----ILGIESSGSVTLTATTEGALAVNSIGSGTVTVTANSGALLTLTAGSTIKGT	1221
OY	564	WKLAMTGTGNKRYG-----EFFMFKINRPPREK-EGNL-----VPIILMGN	603
Db	1222	ESVYTTSSGSGIIGIISGTYEVKATRESLTIOSNKKIKATIGEANVTASGTIGETISGN	1281
OY	604	AVNVRSLMQVOETHAASSLQTDROGLIMIDIGNEFHVASAGEDNIRYRHNSGGYVLSYNNEIT	663
Db	1282	TVNVTYA-----NAGDLT-----VGNCAEINATEGAATLTITSSGKILTEASSHIT	1325
OY	664	PKHTTSAFSQLFSDKRIYAVSNNEYRYALGSLYQYTTSLGNITRRYASRPNVNVGIL	722
Db	1326	-----SAKQVNLISADODGSVASISINAAVTL-----NTGTGLTVTKGSMNATISGTL	1372

RESULT 10  
US-08-719-641-2

```

1  GENERAL INFORMATION:
2  APPLICANT: Barenkamp, Stephen J
3  TITLE OF INVENTION: High Molecular Weight Surface Proteins
4  TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
5  NUMBER OF SEQUENCES: 10
6  CORRESPONDENCE ADDRESS:
7

```

REEDBELL, JOSEPHINE and MARYCATH, MRS.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
;

```

```

; SOFTWARE: Patentin Release #1.0,
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
;

```

```

; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
;

```

```

; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-6255

```

```

; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
MOLECULE TYPE:

```

Query Match	4.08;	Score 192;	DB 4;	Length 1536;
Best Local Similarity	20.7%;	Pred. No. 2.5e-07;		
Matches 174;	Conservative 119;	Mismatches 276;	Indels 270;	Gaps 39

```

QY 12 FSTLLSVDFDLSLATTISLPEDFSGDSQDNARSRNVQAG----- 53
Db 676 FNLT-----DGRGDSACTLTQPIYNLKGISFNKDTTNNVENARVNFEDIKAPIGINKYSS 733
QY 54 -DVSLLTGDSVSI--NYDNSAL-----KKACFNVTSGSVTEAGNHGILYFN 96
Db 732 LMYASFNNISNISVGGSGSVDFLLASSSNVQPGVINYKSFVENSTGS-----SLRFK 783
QY 97 NISSGTTKEGAVILCCODPQATARSGFS-TLSFTQSDEKIDEGCLYSKNAMLLNYYV 155
Db 784 --TSGSTGTGSI---EKDLTLMTGCGITLTYGVEGDMGKGIKAVAKKNITFEGCGNITE 838
QY 156 RFQDNQSKTRKCAL--SGANTVIYG-NYDSVFYQNAATP-GGAIHSSGPQLQIAVNAQAEI 2111
Db 839 GSRRAVEITENVTINNANVTLLIGSDPFDN--HOKPLTLTKDDVIINSGLTYAGAGNIVNI 895

```

QY 212 RFAQNTAKNGSGALYSDDGIDIDONAVLRENEALT--AIGKGAVCCLPTSGSSTPV 270  
Db 896 A-----GILYVESNA-----NFKALTFTFNVG----- 919  
QY 271 PIYFSDNKOLVERHNSINGGAIV-----ARKLSI-----SSGGPLTFINNISYANSQ- 320  
Db 920 -----LFDNKG-----NSNISIANKGARFKDIDNSKLSITTTSSSTYRIIISGNTNNKNDL 972  
QY 321 ---NLGALAIIDGGEIS-----ISAEGTTFQ-----GNFTSLPFLNGHLIQ 362  
Db 973 NITNMGSDTFMOIGQGVDSQKGNLITSSDKINTTKQITIRAGVGENSDADANNALTLI 1032  
QY 363 NAFELKIQANNGSIEFYDITSEADSGTOLNIDGPKNEYT---GTILFSEKSLAND 419  
Db 1033 KTELKLT-----TQDLNITSGFNK-AETITAKDSGLTLTGNTNSADG 1071  
QY 420 PRDFKSTIPQ--NVNLSAGYLVIRKEGAELVTSKFTQSPGSHLYLDLQTKLASKEDIATIT 477  
Db 1072 TNARKVTYFNQVKDSKISA-----DGHKVTLHKSVEITSGSN-----NMTEDSSDN 1115  
QY 478 GLAIDID-----SLSSSTAAYIKANTANKOISVTDSTIELISPGNAYEDL 523  
Db 1116 NAGLITDAKNVTNNNITSHKAVSISATSGEITTKGTITATGAVELTAQTS----- 1170  
QY 524 RMNRSOTFPLLSLEPGAGSVTYTA-----GDFLPVSPHYG-----FGN 563  
Db 1171 -----ILGGIESSGSSVTLTATEGALAVNISGNTVTVANSGLATTLAGSTIKGT 1221  
QY 564 MKLAWTGTGNKRVG-----EFFWDKINYPKRPK-EGNL-----VPNLMGN 603  
Db 1222 ESYTSSQSDIGTIGTVEYKATESLTQJNSKIKATGTGANVTASGTIGTISGN 1281  
QY 604 AVNVRSLMOVQEHASLQDRCLMIDIGCNFPHVASSEDNIRHNSGGYVLSVNNET 663  
Db 1282 TVNVTAT-----NAGDLT-----VNGCAINATGCAATLTSSGKLTTEASHIT 1325  
QY 664 PKHYTSAFQSLSPRDKDVAVSNNEYRMVLSYLYOTTSIGNIFRASNPNVNVGIL 722  
Db 1326 ----SAGOVNLSAODGSVAGSINMANVTL-----NTGTTLTYVKGSNINATSGTL 1372  
RESULT 11  
US-08-728-470-10  
Sequence 10. Application US/08728470  
Patent No. 5928651  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,470  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA: GB 9205704.1  
APPLICATION NUMBER: 16-MAR-1992  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1529 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-728-470-10  
Query Match 3.7%; Score 178.5; DB 2; Length 1529;  
Best Local Similarity 21.2%; Pred. No. 3.7e-06;  
Matches 170; Conservative 109; Mismatches 266; Indels 255; Gaps 40;  
QY 21 DTSLATTSLSLT---PEDSFHGD---SQNAERSYNNQAGDV-----YSLTGVDYSIS 65  
Db 724 DLTINATGNSFSLKQTKRDSFYNEYSKHAINSHTLTLGAVTLGGENSSSTGTGINIT 783  
QY 66 NVNLSALNKACFN-----VTSQSVTFAGN-----HHGLYFN-NISSGTTKEGAVL 109  
Db 784 NKANVTLQADTSNNSNGLKRRITLTLGNSIVEGNLSLGMANAVGNLSIADSTFKGE-- 841  
QY 110 CCDDPQATARFSGFSTLSPFQSPGDIKEGCLYSKANMLNNYVRFEDNOSKTKGAI 169  
Db 842 ASDNLNITGTFNNGANINIKOGVVKLQDINNKGGLNITTT-----NAGTOKTII 893  
QY 170 SG-----ANVTIVGN-----DSVEFYQ-----NAATFGGAIH 197  
Db 894 NGNITTEKGLNINIKAKADEIOTGNGISQKEGNLTSSDKVITNOITTKACVEGGRSD 933  
QY 198 SS-----GPLQIA-VNQAELRFAONTARKNS-----GGALYSPGDIDID 235  
Db 954 SSEAEANANLTIQTKELKLAGDLNISGFNKAEI-----TAKNGSDLTIGNA--SGGNADAK 1006  
QY 236 QNAVYLFRENEALTTAIGKGAVCCLPTSGSSTPVPIYFSDNKOLVERHNSIMG--- 291  
Db 1007 KYTFEDYKDSK-----ISTDGHNVTL-----NSEVYTSNGSSNAGDMS 1045  
QY 292 -GGATYARKSISSGGPTLFINNISYA--NSQNLGALAIIDTGGELISLAEKTTIFQGN 348  
Db 1046 TGLTISAKDVTANNVTSKHTINISAAAGVTTKEGTTINATGSEVAVANONGTI--KGN 1103  
QY 349 RTSLPFLNGIHLILONAKFLKIQARNGCSIEFYDITSEADSGTOLNIDGPKRREYGTI 408  
Db 1104 ITS-----QNTVYATE-----NLVYTE-----NAVINAT- 1128  
QY 409 LBSGEKSLANDPRDFKSTIPQ---NVNLSAGYLVIR---EGAEVYVS---KFTQSPGS 457  
Db 1129 --SGTVNISSTKTDIGKIGESTSGNVTIFASGNTLKYNSNITGDVYVTAADAGALTLAG 1186  
QY 458 HVLVDGTRILASKE-----DIATGLAIDISLSSSTAIVARA---NF 499  
Db 1187 TTSATYGNANITTKTDINDINGKEVSSGSVTLVATGATLAVNISGNTVTTADSGKLTST 1246  
QY 500 ANKOISVTDSTIELISPTGN-----AYEDLRMNSOTF-----PLLSLEPG 539  
Db 1247 VGSTINGTNSVTSQSGDIEGTISGNTYVAVTASGTDLTIGNSAKAEKAGATTLAESG 1306  
QY 540 ----AGGSVYTVAGDPLVPYSPHYGQGMKLA-----WTGCKNKGEEFFWDKINTKPRP 589  
Db 1307 KLTTQFGSSITSSNGQTTTLAKDSSITAGNINANAVTLNLTGTLTTGD---SKIN----- 1358  
QY 590 EKEGNLVPMI-----LWGNVNVNRSILMOVQETHASLQDRCLMIDIGNFFHVASSEDNI 645

Db 1359 ATSGTILINAKDAKLDGAASGDRV--VNATNMS-----GSGNTAKTSSSVNI 1405  
 QY 646 RYRHN--SGGYLSVNNETT 663  
 Db 1406 TGDLTNTINGLNIISENGRT 1425

RESULT 12  
 US-08-719-641-10  
 ; Sequence 10, Application US/08719641  
 ; Patent No. 6218141

GENERAL INFORMATION:  
 APPLICANT: Barenkamp, Stephen J  
 TITLE OF INVENTION: High Molecular Weight Surface Proteins  
 TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.  
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 STREET: Bldg. 1

CITY: Arlington  
 STATE: Virginia

COUNTRY: U.S.A.  
 ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/719,641

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-625

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1529 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-719-641-10

Query Match 3.7%; Score 178.5; DB 4; Length 1529;

Best Local Similarity 21.2%; Pred. No. 3.7e-06;

Matches 170; Conservative 109; Mismatches 266; Indels 255; Gaps 40;

QY 21 DTSLATTSLSLT---PEDEPHGD---SONAERSYVWQAGDV-----YSLTGVSVIS 65

Db 724 DLTINATGNSFSLKQTKDSFYNEYSKHAINSSHLITLIGVNTLGGENSSSITGININIT 783

QY 66 NVDSALNACFN-----VTSQSVTFAGN-----HHGLYFN-NISSGTTKEGAVL 109

Db 784 MNANVTLDADTSNNTGLKRTLTFLGNISVEGNTSLTGANANIVGLSLAEDSTFGE-- 841

QY 110 CCDDPQATRFSGFSLSTIOSPGDIKEGCLYSKNAALMLNNYVVRFEONOSKTSVGA 169

Db 842 ASDNLNITFTNNGTANINIKOGVYKLOGDINNKGGLNITT-----NASGTOKTII 893

QY 170 SG-----ANVTIGNY-----DSVEFYQ-----NAATFGCAIH 197

Db 894 NGNITNEKGDNLINKIKADAEIOIGNISKEGNLITSSDKVITTOITIKACVBEGRSD 953

QY 198 SS-----GPIQIA--VNQAEIRFPAONTAKNGS-----GGALYSDGDIDID 235

Db 954 SSEAENANLTIQTKELKLAGDNLISGFNKAEI-----TAKNGSDLTIGNA--SGNADAK 1006

QY 236 QNAYLFRNEALTTAIGKGAVCCPTSSSTPVPVTFPSDKOLVFERHNSHIMG----- 291

Db 1007 KYTFEDVKDSEK-----ISTDGHVNTL-----NSEKTSNGSSMAONDNS 1045

QY 292 -GGATYARKLSISSGGPTLEINNISA--NSQMLGAIADTGCETLSAERKTIFFOGN 348

Db 1046 TGLTISKADVTYNNNNTVSHKTTINISAAAGNVTKKEGTTINATIGSEVYRQNTI--KGN 1103

QY 349 RTSLPLNLINGIHLQNAKFLKLOARNGCSIEFYDPTISEADGSTOLNINIGDPKREYTG 408

Db 1104 ITS-----QNTVYATE-----NLVYTE-----NAVINAT-- 1128

QY 409 LFSGEKSLANDPRDFKSTIPQ---NVLNAGYLVK---EGAEYVS---KFTQSPGS 457

Db 1129 --SGTVNISFKTGIDIGIESTSGNVNITASGNTLVNSITGVDVTVTADAGALTITAGS 1186

QY 458 HWLDLGTCLIAKSE-----DIATGLAIDISLSSSTFAVIAK---NT 499

Db 1187 TISATTGNANITTKTGIDINGKVESSSGSVTLVATGATLVAGNISGNTVITADSGULTS 1246

QY 500 ANKOISVTDSEILISPTGN-----AYEDLRNRSQTF-----PLSLERP 539

Db 1247 VGSTINGTNSVTSSQSGDEGTISGNTVNVTSQDGLTIGNSAKVEAKNGAATLTAESG 1306

QY 540 ----AGSVTVTAGDPLPSPHYGFGNKKLA-----MTGCKNKGEFPRMDKINKRPR 589

Db 1307 KLTQTGSSSTISSGQTTVTAQDSSAGNINANVTLNTGTLTTGGD--SKIN----- 1358

QY 590 EKEGNLVPNI---LWGNVNVRSLSMQVOETHASSIQTRGMLIDIGNEFFHVSASEDNI 645

Db 1359 ATSGTILINAKDAKLDGAASGDRV--VNATNMS-----GSGNTAKTSSSVNI 1405

QY 646 RYRHN--SGGYLSVNNETT 663

Db 1406 TGDLTNTINGLNIISENGRT 1425

RESULT 13

US-08-038-682-4

; Sequence 4, Application US/08038682

; Patent No. 5549897

GENERAL INFORMATION:

APPLICANT: BARENKAMP, STEPHEN J

TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/038,682

FILING DATE: 16-MAR-1993

```

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-4

Query Match      3.5%; Score 170.5; DB 1; Length 1477;
Best local Similarity 22.4%; Pred. No. 1.7e-05;
Matches 170; Conservative 88; Mismatches 283; Indels 219; Gaps 42;

QY 21 DTSLSAT---TISLPEDSFHDSQNAERS-YNVQ---AGDV-----YSLTGDVSYIS 65
DB 795 DLTINATNSNFSLRQTKDFDYDGYARNAINSTYINISILGNAVTLGGONSSSITGNITIE 854
QY 66 NVNSALN-----KACFNVTSGSVTFAGNHGGLYFN-NISSGTTKEGAVL 109
DB 855 KANVTLEANNAPNOONIRRVKILGSLVNGSLSLGENADIKGNLTJISEATFFKGT- 913
QY 110 CCODP--QATARFSGFSTLSF-----IOSPEDIEQGCLYSKNAFMLNNVVFPEONQSK 163
DB 914 --RDLTINIGNFNNNGTAELINITOGVYKLGAVYNDGDL-----NITTAKRQORS 961
QY 164 TKGAI---SGANVTIVGYDSVSYQNAATFGAI-HSSGPIQIAVNOAEIRFQONTA 218
DB 962 IIGGDIINKKGSINTIDSNDAEIQ-----IGNISQKEGMLTSSDKINIT-KQITI 1013
QY 219 KNSGGALYSDGDIIDONAVYLFRENEALTTAIGKGAVCCPTSGSSTPVIVFSDN 278
DB 1014 KKGIDG---EDSSDANTSNNMLTKIKELKLT-----DISISGFNAE--ITAKDG 1060
QY 279 KOVFERNHISMGGAIVYARKLSISSGPTLFINNISYANSQNLGAIAIDTGGESLSA 338
DB 1061 RDLTIGNSND--GNSGAER-----TVTNNVK-----DKTISAD-CHNVTLNS 1101
QY 339 EKGTTIFQGNRTSLPPLNGIHLQNAKFLKQARNGCSIEFYDPTISEADGSTQLNINCD 398
DB 1102 KVTSSNGRES-----NS-----DNDTGLITAKN--VEVNKDITS---LKTVNTAS 1146
QY 399 PKKKEYTGTL--FSGEKSLANDPRDKSTIPON-VNLSAGYLVIKEGAEVTVSKFTQSP 455
DB 1147 EKVTTAGSTINATNGKASITTKGTDISGTTISGNTVSVA-----TVDLTKFS- 1194
QY 456 GSHLVLDLGTKLASKEDIAITGLAIDISLSSSTAAYIKANTAN-----KOISV 506
DB 1195 -----GSKIAKSGEANVTISATGTIGTISGNTVNV-----TANAGDLTVNGAEIINA 1242
QY 507 TDSIELISPTGNAIEDLRMNSQTFPLSLPEPAGGSVTVTAGDPLVPSPHYFGQ----- 562
DB 1243 TEGAAVLTATGN-----TLTTEAGSSITSTKQOVLLAQNGSIAGSINA 1286
QY 563 -NNKLAWTG---GNKYGFEFWDKINKPRPEKGCNLVPI-----LMG-----N 603
DB 1287 ANVTLNTGTGLTTVAASDI-----KATSGTLVINKAKDKLNGDAGSDSTEVA 1333
QY 604 AVNVRSLMQOETHASS-----LOTDRGLMI---DGIGNEFHVASSEDRIIRRHNSGV 654
DB 1334 AVNAGSGSTATATSSSVNTGDLNVTNGINLTISKDG-RRTVRLRKGELEVKYIQG--- 1389
QY 655 VLSVNNETPPKHYTSMASFOLFSDRDYAVASNNERYMILG 694
DB 1390 VASVEVYEIAK-----RVLEKVKDLSDEREETLAKIG 1421

```

```

RESULT 14
US-08-302-832-4
Sequence 4, Application US/08302832
Patent No. 5603938
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: Of No. 5603938-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-302-832-4

Query Match      3.5%; Score 170.5; DB 1; Length 1477;
Best local Similarity 22.4%; Pred. No. 1.7e-05;
Matches 170; Conservative 88; Mismatches 283; Indels 219; Gaps 42;

QY 21 DTSLSAT---TISLPEDSFHDSQNAERS-YNVQ---AGDV-----YSLTGDVSYIS 65
DB 795 DLTINATNSNFSLRQTKDFDYDGYARNAINSTYINISILGNAVTLGGONSSSITGNITIE 854
QY 66 NVNSALN-----KACFNVTSGSVTFAGNHGGLYFN-NISSGTTKEGAVL 109
DB 855 KANVTLEANNAPNOONIRRVKILGSLVNGSLSLGENADIKGNLTJISEATFFKGT- 913
QY 110 CCODP--QATARFSGFSTLSF-----IOSPEDIEQGCLYSKNAFMLNNVVFPEONQSK 163
DB 914 --RDLTINIGNFNNNGTAELINITOGVYKLGAVYNDGDL-----NITTAKRQORS 961
QY 164 TKGAI---SGANVTIVGYDSVSYQNAATFGAI-HSSGPIQIAVNOAEIRFQONTA 218
DB 962 IIGGDIINKKGSINTIDSNDAEIQ-----IGNISQKEGMLTSSDKINIT-KQITI 1013
QY 219 KNSGGALYSDGDIIDONAVYLFRENEALTTAIGKGAVCCPTSGSSTPVIVFSDN 278
DB 1014 KKGIDG---EDSSDANTSNNMLTKIKELKLT-----DISISGFNAE--ITAKDG 1060

```

```

OY 279 KOLVERNRHSIMGCAIYARKLSISSGGPFLFINNISYANSONMIGALAIIDGGEISLSA 338
DB 1061 RDLTIGNSND--GNSGAERK-----TVTGNNVK-----DSKISAD--GHNTVITNS 1101
OY 339 EKGTTFQGNRSLPFLNGIHLQNAKFLKQARNCGSIEFYDPTISEADGSTQOLINCD 398
DB 1102 KVTSSSNGGRES-----NS-----DNDTGLITITAKN--VEVNRDITS---LKTVINITAS 1146
OY 399 PKRKEVGTTL--FSGEKSLANDPRDEKSTIPON--VNLSGYLVIREGAEVTSKFTQSP 455
DB 1147 EKVTTAGSTINATNGKASITTKGTGDISGTSIGNTVSVSA-----TYVDLTTKS- 1194
OY 456 GSHLVLDLGTKLIAKEDIAITGLAIDISLSSSTAAYIKANTAN-----KOISV 506
DB 1195 -----GSKIKAKEGAEANVTATGTTIGTIGTIGNTVN-----TANAGDLTVNGAEIINA 1242
OY 507 TDSIELISPTGNAYEDLRMRNSOTFPLLSLEPGAGSVTVTAGDFLPVSPHYFGQ---- 562
DB 1243 TEGAAITLITATGN-----TLTTEAGSSITSTKQVLDLQNGSIASINA 1286
OY 563 -NMKLAWTGT-----GNKYGEFFWMDKINPKRPEKEGNLVPNL-----LWG-----N 603
DB 1287 ANVTLMVTGTLTVAGSDI-----KATSGFLVINKDAKLANGDASGDESTEVN 1333
OY 604 AVNVRSLMOYQETHASS-----LQTRDGLMI---DGIGNFFHVSASEDNIRYRHNSGVY 654
DB 1334 AVNASSGSVTAATSSSVNTGDLNVTNGINLITSKDG--RTVRLRKELEVKYIQG--- 1389
OY 655 VLSVNNETPKHTYSMAFSOLFSDRDYAVSNNEYRYMYS 694
DB 1390 VASVEEVIETAK-----RVLEKVKDLSDEERETLAKLG 1421

RESULT 15
US-08-530-198-4
; Sequence 4, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-530-198-4

Query Match 3.5%; Score 170.5; DB 2; Length 1477;
Best Local Similarity 22.4%; Pred. No. 1.7e-05;
Matches 170; Conservative 88; Mismatches 283; Indels 219; Gaps 42;

OY 21 DTSLSAT-----TSLPDESFHGDSONAERS--YNYO--AGDV-----YSLTGVSVIS 65
DB 795 DLTINATNSFSLRQTKDDPYDGYARALNSTYINISLIGNVTLGGONSSSITGNTTIE 854
OY 66 NYDNSALN-----KACFNVTSGSVTPAGNHGLYFN--NISSGTTKEGAVI 109
DB 855 KAAVNTLEANNAPNOQINRDIRVILKISLVNGSLSLTGERADIKHMLTISESATFPGKT- 913
OY 110 CCODP-QATARESGFSLSE-----IQSPEDKEQGLSKNALMLLNTVVFEDQNSK 163
DB 914 --RDTLINTGNFNNNGTAEINITQGVYKLGNTYNDGL-----NITTHAKNORS 961
OY 164 TRGGAIT-----SGAVNTIVGNDVSFYNAAATFGAI--HSSGPLOJAVNOAEIRFQONTA 218
DB 962 IIGGDIINKKGSINTIDSNDAEIQ-----IGNISOKEGNLTJSSDKINIT- KOITI 1013
OY 219 KNGSGALYSDGDIDIDONAYVLFRENEALTTAIGKGAVCCLPTSGSSTPVPIVTFPSDN 278
DB 1014 KRKIDG---EDSSSDATSNANLTKIKELKLT-----DLISGFMKAE--ITAKDG 1060
OY 279 KOLVERNRHSIMGCAIYARKLSISSGGPFLFINNISYANSONMIGALAIIDGGEISLSA 338
DB 1061 RDLTIGNSND--GNSGAERK-----TVTGNNVK-----DSKISAD--GHNTVITNS 1101
OY 339 EKGTTFQGNRSLPFLNGIHLQNAKFLKQARNCGSIEFYDPTISEADGSTQOLINCD 398
DB 1102 KVTSSSNGGRES-----NS-----DNDTGLITITAKN--VEVNRDITS---LKTVINITAS 1146
OY 399 PKRKEVGTTL--FSGEKSLANDPRDEKSTIPON--VNLSGYLVIREGAEVTSKFTQSP 455
DB 1147 EKVTTAGSTINATNGKASITTKGTGDISGTSIGNTVSVSA-----TYVDLTTKS- 1194
OY 456 GSHLVLDLGTKLIAKEDIAITGLAIDISLSSSTAAYIKANTAN-----KOISV 506
DB 1195 -----GSKIKAKEGAEANVTATGTTIGTIGTIGNTVN-----TANAGDLTVNGAEIINA 1242
OY 507 TDSIELISPTGNAYEDLRMRNSOTFPLLSLEPGAGSVTVTAGDFLPVSPHYFGQ---- 562
DB 1243 TEGAAITLITATGN-----TLTTEAGSSITSTKQVLDLQNGSIASINA 1286
OY 563 -NMKLAWTGT-----GNKYGEFFWMDKINPKRPEKEGNLVPNL-----LWG-----N 603
DB 1287 ANVTLMVTGTLTVAGSDI-----KATSGFLVINKDAKLANGDASGDESTEVN 1333
OY 604 AVNVRSLMOYQETHASS-----LQTRDGLMI---DGIGNFFHVSASEDNIRYRHNSGVY 654
DB 1334 AVNASSGSVTAATSSSVNTGDLNVTNGINLITSKDG--RTVRLRKELEVKYIQG--- 1389
OY 655 VLSVNNETPKHTYSMAFSOLFSDRDYAVSNNEYRYMYS 694
DB 1390 VASVEEVIETAK-----RVLEKVKDLSDEERETLAKLG 1421

```

Search completed: October 2, 2001, 03:27:36  
Job time: 7218 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2001, 03:28:49 ; Search time 76.79 Seconds  
(without alignments)  
914.610 Million cell updates/sec

Title: US-09-446-677B-6

Perfect score: 4814

Sequence: 1 MRFSLGCPPLVFSLTLSVF.....IECRPHANVINGSKRRF 922

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4794	99.6	922	2	B72131 polymorphic outer
2	4794	99.6	922	2	E86491 polymorphic outer
3	4790	99.5	922	2	F81539 polymorphic membra
4	1441	29.9	928	2	E86546 polymorphic outer
5	1441	29.9	928	2	B72077 polymorphic membra
6	1434	29.8	930	2	D86546 polymorphic outer
7	1434	29.8	930	2	A81591 polymorphic membra
8	1432	29.7	930	2	D72078 polymorphic outer.
9	1400.5	29.1	936	2	C86546 polymorphic membra
10	1400.5	29.1	936	2	B81591 polymorphic membra
11	1400.5	29.1	936	2	C72078 polymorphic outer
12	1393	28.9	928	2	G86546 polymorphic outer
13	1393	28.9	928	2	G81591 polymorphic membra
14	1344	27.9	928	2	H86546 polymorphic outer
15	1344	27.9	928	2	D72077 polymorphic membra
16	1344	27.9	949	2	F81591 polymorphic membra
17	1249	25.9	772	2	H86492 Pmp_3 [imported] -
18	1137.5	23.6	973	2	B86547 polymorphic outer
19	1137.5	23.6	973	2	F72076 polymorphic membra
20	1137.5	23.6	995	2	C81593 polymorphic outer
21	1060.5	22.1	1407	2	B72078 polymorphic membra
22	1060	22.0	1276	2	B86546 polymorphic outer
23	1060	22.0	1276	2	C81591 polymorphic membra
24	1007	20.9	841	2	E72130 polymorphic membra
25	951.5	19.8	1013	2	G71460 probable outer mem
26	901	18.7	987	2	H81722 polymorphic membra
27	796.5	16.5	712	2	E86492 polymorphic outer
28	774.5	16.1	867	2	F81721 polymorphic membra
29	727	15.1	878	2	B71460 probable outer mem

30	720	15.0	445	2	E86493 Pmp_5 [imported] -
31	629.5	13.1	359	2	C86493 Pmp_4 [imported] -
32	610	12.7	1723	2	H86557 polymorphic membra
33	610	12.7	1723	2	E72067 polymorphic membra
34	610	12.7	1732	2	C81601 polymorphic membra
35	581	12.1	946	2	C86549 polymorphic outer
36	581	12.1	946	2	D81594 polymorphic membra
37	581	12.1	946	2	C72075 polymorphic outer
38	577	12.0	1609	2	A86611 probable outer mem
39	577	12.0	1609	2	H72013 polymorphic membra
40	576.5	12.0	427	2	A86493 polymorphic outer
41	574	11.9	938	2	F86548 polymorphic membra
42	574	11.9	938	2	H72074 polymorphic membra
43	569.5	11.8	934	2	G86548 polymorphic outer
44	569.5	11.8	934	2	A72075 polymorphic outer
45	569.5	11.8	952	2	D81593 polymorphic membra

## ALIGNMENTS

RESULT 1  
B72131 polymorphic outer membrane protein g family - Chlamydomonas pneumoniae (strain CWL029)  
C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: B72131  
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999  
A>Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000, NCID:99206606  
A:Accession: B72131  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-922 <ARN>  
A:Cross-references: GB:AE001585; GB:AE001363; NID:94376255; PID:AA018163.1; PID:9437  
A:Experimental source: strain CWL029  
C:Genetics:  
A:gene: pmp\_1

Query Match	Score	DB 2;	Length	922;
Best local Similarity	99.6%;			
Matches	919;	Conservative	1;	Mismatches
			2;	Indels
			0;	Gaps
			0;	
QY	1	MRFSLGCPPLVFSLTLSVFDTLSATTSVLPEDSFGDSQNAERSYNVQAGDVSITG	60	
DB	1	MRFSLGCPPLVFSLTLSVFDTLSATTSVLPEDSFGDSQNAERSYNVQAGDVSITG	60	
QY	61	DVSISVNDNSALKACPNVSGSVTFAGNHGLEYFNISGTTREGAVLCCODQATARF	120	
DB	61	DVSISVNDNSALKACPNVSGSVTFAGNHGLEYFNISGTTREGAVLCCODQATARF	120	
QY	121	SGFSTLSFIOSPGDIKROGLYSKNAIMLNNTVVRQDSKTKGASGANTYVGN	180	
DB	121	SGFSTLSFIOSPGDIKROGLYSKNAIMLNNTVVRQDSKTKGASGANTYVGN	180	
QY	181	DVSFTFONATFGGAIHSSGPDQIAVNOAEIRFQNTAKGSGGALYSDGDDIDQNAV	240	
DB	181	DVSFTFONATFGGAIHSSGPDQIAVNOAEIRFQNTAKGSGGALYSDGDDIDQNAV	240	
QY	241	LFRENALTTAIGKGVCCLPISGSSTPVPIVTFESNKOVLFEERNHNSIMGGAIVARL	300	
DB	241	LFRENALTTAIGKGVCCLPISGSSTPVPIVTFESNKOVLFEERNHNSIMGGAIVARL	300	
QY	301	SISGSPTLFNNISYANSQNLGATAIDTGEISLSAEKGTTFPGNRTSLPLNGIHL	360	
DB	301	SISGSPTLFNNISYANSQNLGATAIDTGEISLSAEKGTTFPGNRTSLPLNGIHL	360	
QY	361	LQNAKFLKLOARNGSIEFYDPTTSEADGSTQNLINDPKNKRYTGTILFSGKSLAND	420	
DB	361	LQNAKFLKLOARNGSIEFYDPTTSEADGSTQNLINDPKNKRYTGTILFSGKSLAND	420	

Qy 421 RDKFKTIQNNNLNSGLVYIEKEAEVTVSKFTQSGSHLVYDLGKTLASKEDIAITGLA 480  
Db 421 RDKFKTIQNNNLNSGLVYIEKEAEVTVSKFTQSGSHLVYDLGKTLASKEDIAITGLA 480  
Qy 481 IDIDLSSSSTAAVYIKANTANKQISVTSIELISPTGNAEDELRRMNSQTPPLLSLEPGA 540  
Db 481 IDIDLSSSSTAAVYIKANTANKQISVTSIELISPTGNAEDELRRMNSQTPPLLSLEPGA 540  
Qy 541 GGSVTVTAGDELVPSPHYGFOGNNKLAWTGTGNKYGEFFWDKINYPKPEKEGNLVNPL 600  
Db 541 GGSVTVTAGDELVPSPHYGFOGNNKLAWTGTGNKYGEFFWDKINYPKPEKEGNLVNPL 600  
Qy 601 WGNNAVVSILMOVOVTHASSILOTDRGLWIDIGNEFHHVASBDNRIRHNSGGYLVSVNN 660  
Db 601 WGNNAVVSILMOVOVTHASSILOTDRGLWIDIGNEFHHVASBDNRIRHNSGGYLVSVNN 660  
Qy 661 EITPKHTYSMAFSQLESRDKYAVSNNEYRMYLGSYLYQYTTSLGNIFRYASRPNVNWG 720  
Db 661 EITPKHTYSMAFSQLESRDKYAVSNNEYRMYLGSYLYQYTTSLGNIFRYASRPNVNWG 720  
Qy 721 ILSRRELQNPMLIFELFLAYGHATIDMKTDYANPPMYNNSRRNNCMALIECGSNPLVFE 780  
Db 721 ILSRRELQNPMLIFELFLAYGHATIDMKTDYANPPMYNNSRRNNCMALIECGSNPLVFE 780  
Qy 781 NGRLEFGAIPFMKLOLVYAYOGDEFKETTADGRFRFSGNSLTGISVPLGIRFEKIALSDOVL 840  
Db 781 NGRLEFGAIPFMKLOLVYAYOGDEFKETTADGRFRFSGNSLTGISVPLGIRFEKIALSDOVL 840  
Qy 841 YDFSFSTYIPDIFRKPDSCDEALVYISGDSMLVPAHVSHPAFVSGTGRHFNDDYTELLCR 900  
Db 841 YDFSFSTYIPDIFRKPDSCDEALVYISGDSMLVPAHVSHPAFVSGTGRHFNDDYTELLCR 900  
Qy 901 GSTICRPHARNYNNINGCKSKFRF 922  
Db 901 GSTICRPHARNYNNINGCKSKFRF 922

RESULT 2  
E86491  
polymorphic outer membrane protein G family [imported] - Chlamydomonas reinhardtii (strain  
C:Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: E86491  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Uuchi, K.; Shiba, T.; Ishi-  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydomonas reinhardtii J138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: E86491  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-922 <SNO>  
A:Cross-references: CB:BA000008; NID:98978378; PIDN:BA098215.1; GSPDB:GN00142  
A:Experimental source: strain J138  
A:Genetics:  
A:Gene: pmp-1

Query Match	99.6%	Score 4794	DB 2	Length 922
Best Local Similarity	99.7%	Pred. No. 2e-281		
Matches 919	Conservative 1	Mismatches 2	Indels 0	Gaps 0
Qy	1	MRFSLCGPLVFLSLTSLVFDTSLSATTSLTLPEDSFHGDSQNAERSTNYVAGVYSLTG	60	
Db	1	MRFSLCGPLVFLSLTSLVFDTSLSATTSLTLPEDSFHGDSQNAERSTNYVAGVYSLTG	60	
Qy	61	DVLSINVDNSALNKAFCENVTSGSVTFAGNHHGLYFNNISSGTTEGAVLCCDDQDQATARE	120	
Db	61	DVLSINVDNSALNKAFCENVTSGSVTFAGNHHGLYFNNISSGTTEGAVLCCDDQDQATARE	120	
Qy	121	SGEFTLSLTGSPGDIKEGGCLYSKNAIMLNNVYVREFQDNQSKTKGGAISGANTTYGNT	180	
Db	121	SGEFTLSLTGSPGDIKEGGCLYSKNAIMLNNVYVREFQDNQSKTKGGAISGANTTYGNT	180	

QY	181	DSVSYQONNATFGAIHSSGPLQIAVNOAEIRFQANTAKNSGGLVSDGDIIDONAVY	240
Db	181	DSVSYQONNATFGAIHSSGPLQIAVNOAEIRFQANTAKNSGGLVSDGDIIDONAVY	240
QY	241	LFRENEALTTAIGKGAACVCLPTSGSSTPVPIVTFESDKOLVFEENHISIMGGAIFYARKL	300
Db	241	LFRENEALTTAIGKGAACVCLPTSGSSTPVPIVTFESDKOLVFEENHISIMGGAIFYARKL	300
QY	301	SISGGPTLFINNISYANSQNLGAIADTGGEISLSAEKGTTFQGNRTSLPFLNGIHL	360
Db	301	SISGGPTLFINNISYANSQNLGAIADTGGEISLSAEKGTTFQGNRTSLPFLNGIHL	360
QY	361	LONAFPLKQANNGSIEFYDPITSEADGSOVLNINBPKKREYTGTLFEGEKSILANDP	420
Db	361	LONAFPLKQANNGSIEFYDPITSEADGSOVLNINBPKKREYTGTLFEGEKSILANDP	420
QY	421	RDFKSTIQONVLSGXYLVIRKEGAEVTSKFTQSPGSHLYDLGTLKILASKEDIAITGLA	480
Db	421	RDFKSTIQONVLSGXYLVIRKEGAEVTSKFTQSPGSHLYDLGTLKILASKEDIAITGLA	480
QY	481	IDIDSLSSSSTAIVIKANTANKQISVTDSEILISPTGNAVEDLRRNRSQTPPLSLERGA	540
Db	481	IDIDSLSSSSTAIVIKANTANKQISVTDSEILISPTGNAVEDLRRNRSQTPPLSLERGA	540
QY	541	GGSVTVTAGDFLPVSPHVGFOGMMKLTAMGTGNKXGEFFMDKINPKPREKGNLVPIIL	600
Db	541	GGSVTVTAGDFLPVSPHVGFOGMMKLTAMGTGNKXGEFFMDKINPKPREKGNLVPIIL	600
QY	601	WGNNAVNSLMOVQETHASSLOTDRGLWIDIGNFHYSASBDNIRYHNSGGYVLVYNN	660
Db	601	WGNNAVNSLMOVQETHASSLOTDRGLWIDIGNFHYSASBDNIRYHNSGGYVLVYNN	660
QY	661	ETTPKHVYTSMAFSQLEFSRDKDYAVSNNEYRMVLSGLYQYTTSLGNIIRYASRBNPNVNG	720
Db	661	ETTPKHVYTSMAFSQLEFSRDKDYAVSNNEYRMVLSGLYQYTTSLGNIIRYASRBNPNVNG	720
QY	721	ILSRFQLONPLMIHFELCAYGHATDMKTVDYANFMYNNSNRNNOMALECGSMPLLYFE	780
Db	721	ILSRFQLONPLMIHFELCAYGHATDMKTVDYANFMYNNSNRNNOMALECGSMPLLYFE	780
QY	781	NGRLQOGAIPFVKLOLVAYOGDFEKETTAADRRFSNGSLTISVPLGIRFEKIALSDOVL	840
Db	781	NGRLQOGAIPFVKLOLVAYOGDFEKETTAADRRFSNGSLTISVPLGIRFEKIALSDOVL	840
QY	841	YDFESYSPIDIFERKDPSCAALVYISGDSMLVPAAHVSHAFVSGGTGRHENDYTELLCR	900
Db	841	YDFESYSPIDIFERKDPSCAALVYISGDSMLVPAAHVSHAFVSGGTGRHENDYTELLCR	900
QY	901	GSIECRPHARNYNINCGSKFRF 922	
Db	901	GSIECRPHARNYNINCGSKFRF 922	

RESULT 3  
F81539  
polymorphic membrane protein G family CP0770 [imported] - Chlamydia pneumoniae (S)  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: F81539  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255  
A:Accession: F81539  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-922 <REA>  
A:Cross-references: GB:AE002237; GB:AE002161; NTD:97189684; PIDN:AAF38570.1; PID:g718  
A:Experimental source: Strain AK39, HL cells  
C:Genetics:  
A:Gene: CP0770

Query Match 99.5%; Score 4790; DB 2; Length 922;  
 Best Local Similarity 99.6%; Pred. No. 3.5e-281;  
 Matches 918; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 MRSFSGPLVFSLLTSLVFDTSLSATTISLTPEDEPHGDSOMASERYNVOAGDVSILG 60
DB 1 MRSFSGPLVFSLLTSLVFDTSLSATTISLTPEDEPHGDSOMASERYNVOAGDVSILG 60
QY 61 DVSISVNDALMKACPRNVTSGSVTPAGNHGLEYFNINSSGTTKEGAVLCCODPQATARF 120
DB 61 DVSISVNDALMKACPRNVTSGSVTPAGNHGLEYFNINSSGTTKEGAVLCCODPQATARF 120
QY 121 SGFSTLSFIOGPDIOGGLYKSKNMLMLNNTVVRFEQNSQTKGGAISGANTYIGNY 180
DB 121 SGFSTLSFIOGPDIOGGLYKSKNMLMLNNTVVRFEQNSQTKGGAISGANTYIGNY 180
QY 181 DVSFYONAAATFGAIIHSSGPIQIAVNOAEIRFQAONTAKNGSGALYSDDIDIDONAVY 240
DB 181 DVSFYONAAATFGAIIHSSGPIQIAVNOAEIRFQAONTAKNGSGALYSDDIDIDONAVY 240
QY 241 LFRFNALTTAICKGAVCCLPYTSSTPVPYTFSDNKOVLERNHISMGGAITYARKL 300
DB 241 LFRFNALTTAICKGAVCCLPYTSSTPVPYTFSDNKOVLERNHISMGGAITYARKL 300
QY 301 SISSGGPTLFINNISYANSONLGATAIDTGGELISAEKGTTFQGNRTSLPELNGIHL 360
DB 301 SISSGGPTLFINNISYANSONLGATAIDTGGELISAEKGTTFQGNRTSLPELNGIHL 360
QY 361 LQNAKFLKQARNGCSIEFYDPTTSEADSGTQJLNDPKNKEYTGTILFSGEKSLANDP 420
DB 361 LQNAKFLKQARNGCSIEFYDPTTSEADSGTQJLNDPKNKEYTGTILFSGEKSLANDP 420
QY 421 RDRKSTIPQAVNLSAGTYLKEGAEVTVSKFTQSPGSHVLDTGTLKILASKEDIAITGLA 480
DB 421 RDRKSTIPQAVNLSAGTYLKEGAEVTVSKFTQSPGSHVLDTGTLKILASKEDIAITGLA 480
QY 481 IDIDSSSSSTAIVIKANTANKOISVTDIELISPTGNAAEDLRMRNSQTFPLLSLEPGA 540
DB 481 IDIDSSSSSTAIVIKANTANKOISVTDIELISPTGNAAEDLRMRNSQTFPLLSLEPGA 540
QY 541 GGSVVTYAGDFLPVSPHYGFGQNMKLTMTGKNGVGEFFMDKINXKPREKEGNLVPNL 600
DB 541 GGSVVTYAGDFLPVSPHYGFGQNMKLTMTGKNGVGEFFMDKINXKPREKEGNLVPNL 600
QY 601 WGAIVNRSLMOVQETHASSLQTDRCGLMDIGIGNFHVASSEDNIRYRHNSGGYVLSVN 660
DB 601 WGAIVNRSLMOVQETHASSLQTDRCGLMDIGIGNFHVASSEDNIRYRHNSGGYVLSVN 660
QY 661 ETRPKHYTSAFSQLSRKQDVAVSNNEHYMGLSYLYQTTSLGNIFFKASRNPNVNG 720
DB 661 ETRPKHYTSAFSQLSRKQDVAVSNNEHYMGLSYLYQTTSLGNIFFKASRNPNVNG 720
QY 721 ILSRRFLQNPMLFHLFCAYGHAINDMKTDYANFPVKNSWRNNCAIEGGSMPLLVFE 780
DB 721 ILSRRFLQNPMLFHLFCAYGHAINDMKTDYANFPVKNSWRNNCAIEGGSMPLLVFE 780
QY 781 NGRFLQGAIPFKLQLVAYVQGDFFKETAADGRRSNGSLTISVPLGIRREKLALSDOVL 840
DB 781 NGRFLQGAIPFKLQLVAYVQGDFFKETAADGRRSNGSLTISVPLGIRREKLALSDOVL 840
QY 841 YDSESEYIDIDIPKQDSCEALVYISGDSWLVPAHAHVSRAHAFVSGRGYHFNDYTELCH 900
DB 841 YDSESEYIDIDIPKQDSCEALVYISGDSWLVPAHAHVSRAHAFVSGRGYHFNDYTELCH 900
QY 901 GSIECRPHANTYINGSKFRF 922
DB 901 GSIECRPHANTYINGSKFRF 922

```

RESULT 4  
 E86546  
 polymorphic outer membrane protein G/I family [imported] - Chlamydia pneumoniae (str

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
 C:Accession: E86546  
 R:Shitai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A:Reference number: A86491, M01D:20330349  
 A:Accession: E86546  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Restrictions: 1-928 <STO>  
 A:Cross-references: GB:BA000008; NID:98978819; PIDN:BA98655.1; GSPDB:GN00142  
 A:Experimental source: strain J138  
 C:Genetics:  
 A:Gene: pmp\_9

Query Match 29.9%; Score 1441; DB 2; Length 928;  
 Best Local Similarity 36.2%; Pred. No. 4.5e-79;  
 Matches 348; Conservative 169; Mismatches 373; Indels 70; Gaps 25;

```

QY 1 MRSFSGPLVFSLLT-LSY-FDTSLSATTISLTPEDEPHGDSOMASERYNVOA-GDVS 57
DB 1 MRSFSGPLVFSLLT-LSY-FDTSLSATTISLTPEDEPHGDSOMASERYNVOA-GDVS 57
QY 58 LTRDVSISVNDN-SALNKACFNTSGSVTPAGNHGLEYFNINSSGTTKEGAVLCCODPQA 116
DB 61 LTRDVSISVNDN-SALNKACFNTSGSVTPAGNHGLEYFNINSSGTTKEGAVLCCODPQA 116
QY 117 TAR---FSGFTLSFIOGPDIOGGLYKSKNMLMLNNTVVRFEQNSQTKGGAISGA 172
DB 114 AANKLSFSGFYSLSLQTNATTTGAIKSTGACSIQSVSCFYGNFENDMGALOGS 173
QY 173 NNTIVGNYSVSFYONAAATFGAIIHSSGPIQIAVNOAEIRFQAONTAKNGSGALYSDD 231
DB 174 SISLSLN-PMLTFKAKNATOKGALYSTGGITINNTLSASFSEMTAAN-NGGAIYFEAS 231
QY 232 IDIDONAVYVFRFNALTTAICKGAVCCLPYTSSTPVPYTFSDNKOVLERNHISMG 291
DB 232 SFISNKAISF-INNSTATATGATGATYC--SSTAPKPVLLSDGEINFTGNATITS 287
QY 292 GGAITYARKLSISSGPTLFINNISYANSONLGATAIDTGGELISAEKGTTFQGN--- 348
DB 288 GGAITYARKLSISSGPTLFINNISYANSONLGATAIDTGGELISAEKGTTFQGN--- 347
QY 349 -----RTSLPPLNGIHLQNAKFLKQARNGCSIEFYDPTTSE--ADSGTQJLND 396
DB 348 KGASSQTTFRNINIGN-----TNAKIYOLRASQGWTFYDPTTSTYALSDALNLM 402
QY 397 GD--PKNKEYTGTILFSGER---SLANDPRDFKSTIPQAVNLSAGTYLKEGAEVTVSK 451
DB 403 GPDLANPAAQGTIVFSGEELSBAEADNLSTIQDPTLLAGGQSLKSGVTLVAKSF 462
QY 452 TOSPGSHVLDTGTLKILASKEDIAITGLAIDIDSSSTAIVIKANTANKOISVTDIE 511
DB 463 SGPSTLLMDACTTL-ETADGTTINNLVNVLSLEKTK-KATLKATQASQTYTLLSLS 520
QY 512 LISPTGNAYEDLRMRNSQTFPLLSLEPGAAGSVTV--AGDFLPVSP-HYFGQNMKLTAV 568
DB 521 LVDPGSNVEDVSWNNPQVFSCLTLADDPANHITDLADPLEKPIHMGYGNMALSW 580
QY 569 ---TGKNGVGEFFMDKINXKPREKEGNLVPNLINGNAVNSLSMOVQETHASSLQTD 625
DB 581 QEDTARKSRAATLTWTKGTGNPBERGTLVANTLWGSFVDSIQDLVATKYROSQETR 640
QY 626 GLMIDGIGNFHVASSEDNIRYRHNSGGYVLSVNNETPRKHYTSAFSQLSRKQDVAVS 685
DB 641 GTWCEGISNFPKQDSCEALVYISGDSWLVPAHAHVSRAHAFVSGRGYHFNDYTELCH 700
QY 686 NNEYRMYLGSYLYQY--TSLGNIFFKASRNPNVNGILSRFLQNPMLFHLFCAYGHA 743
DB 701 KNASAYASLSHLQHLATLSSPILRYLPQSES-----EOP-VLFDAQISYIYS 748

```



```

QY 180 YDSVEFYQNA--TFCGAIHSSGPIQIAVNOAEIRPACQMTAKNGSGCALYSOGDIDIDQNA 238
Db 186 TSSITSTENSASKLIGALYSSAAIASISGNTGOLVFMNNKKEFG-GGALGFEKSSSITONS 244
QY 239 YLIFRENEALPTTAIGKGGAVCCLPPTSGSSFPPIPTTFSDNKOLVERHNSJINGGAIYAR 298
Db 245 SLIFSGNNA-IDAAGKGAICYCKTG----EPTLITISGKNSLTPAEMSSYTOGGAICAH 299
QY 299 KLSISSGGPTLFINNIYSANSQNLGAIADTGCETLSAEKGTITFOGN--RTSLD- 353
Db 300 GLDISAAGPTLFSSNRCGNTAAGKGAIAIADSGSLISANOGDITFLGNTLSTSAPTS 359
QY 354 FLNGHLLQNKFLKQARNCGSTIEFYDPTISEADGSGNQ-LINDDPNK--EYTGTLIF 410
Db 360 TRMALYITLSSSKATINLRAAGQSOTIFYDPIASNTTGAADVLTITNPDNSPLDSSGTYIF 419
QY 411 SGEKSLADPR---DFKSTIPQNVLSAGVLYIKEGAEVTSKFPQSGSHVLDTLGTKL 467
Db 420 SGEKLSADENAAADFTSILKQPLALASGTALAKONVELDNGFTQTIGSTILLMPGKTL 479
QY 468 TASKEDIAITGLAIDIDSLSSSSTAAYIKANTANKOISVTDISIELISPTGNAYEDLRMRN 527
Db 480 KADFEAISLTKLVYDLSALEGNKSVSIEFTAG-ANKTITLTPLVQDSSGNFYESHTINQ 538
QY 528 SOTFPLSLLEG-AGGSYTYVAGDELPY---SPHYGFGQNNKLMATYTCGN-KVGEFFMDK 582
Db 539 AFTQPLVFTATAATASDIYIDLALLTSPQTEPDEPHGYGGMHEATWADSTAKSGMTAVT 598
QY 583 INYKRPEKEGNLVNITLNGNAVNYRSLMOYETFASSLTQDRGLMIDGICNFPVHAS 642
Db 599 TGYNPNPERRASVYDPDSIMASFTDITRTILOQIMTSQANSITYOORGLMAGSTANFFHKDSSG 658
QY 643 DNIRYRHNSSGCVLSSVNNETIPKHYTSMAFSOLFSRDXDYAVSNNEYBMYLGSYLYQYTT 702
Db 659 TNOAFRHKSYGYIYOGSADPSENIFSVAFQOLFGKDXDLFVEMTSHNYLASYLQHRA 718
QY 703 SLGNIFERKASRPNVNVGILSRFLQNPMLTFHFLCAVGAHTNDMKTDYANFPYKNSWR 762
Db 719 FLGGI-----PMPSEFSITDMUKDIP-LILMQLSSTYRKNDMDYRTYSTPEKOGSMT 770
QY 763 NNCVAILIECGSMPLLVFENGRLFOGAIPEMKLOLYAYVAGQFKEKTTAGRRNSGSLSI 822
Db 771 NNSGALIEGGSLALYLPKRAPFPQGYFFPLKFOAVYNSHQOQFKESGAPARAFDGDGLVNC 830
QY 823 SVPLGIREKTLASQDVLYDEFSYSYIPDIFRKDPSCECAALYISGDGMLVPAHYSRHAFF 882
Db 831 SIPGIRLEKISEDEKKNFEISLAIYIGDYVRKNRPSTSLWVSGASWPSLCKNLARQAF 890
QY 883 GSGGGRHYFNQYTELTCGSIEDCRHANYNINGSKRF 922
Db 891 ASAGSHLTLSPHVELSGEAAEYLRGSAHIYVNDGGLRASF 930

```

RESULT 7

A:181591

polymorphic membrane protein G family CP0307 [imported] - Chlamydophila pneumoniae (strain AR39).

C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000

C:Accession: AB1591

R:Read, T.D.; Bringham, R.C.; Shen, C.; Gall, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.L.; et al. (2000) Genome sequence of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39. Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: AB1500; M01D:20150255

A:Accession: AB1591

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-930 <FEAT>

A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AAF38164.1; PID:g71892234

A:Experimental source: strain AR39, HL cells

C:Genetics:

C:Gene: CP0307

Query Match	29.8%;	Score 1434;	DB 2;	Length 930;
Best Local Similarity	36.6%;	Pred. NO. 1.2e-78;		
Matches 344;	Conservative 157;	Mismatches 395;	Indels 44;	Gaps 21;

QY	10	LVFSLTVEFPTS--ATTISLPEPSFH--GSOAENRKYNYAQGVYSLTQDVSIS	65
Db	8	LLISLTVTPILSLATYAGDASLSPDSFDGAGSTTPKSTADANGNTYLSGNYIN	67
QY	66	NV-DNSLANKACENVTSGSVTFAGHHGLTFENNISGTTKEGAVILCODPOATPARGFS	124
Db	68	DAGKGTALGCCFTFTTQDILFTFGKGYSFSEFTVYAGNAGAASTADKALT--FTGGS	125
QY	125	TLSFOSPGDIKEOG--GLYSKNALMLNNYVYRFEQ--NOSKTKGALSGANTYIGN	179
Db	126	NLSFIAAGCTTYASOKSKTSLSSAGALNTDNGIILFSQWVNSNANNNGAIFTAKTISGN	185
QY	180	YDVSFFYONAA--TFGCAIHSSGPLOIAVNOAEIFRPAONTAKNGSGALYSDGDDIDONA	238
Db	186	TSSLFTFNSAKKLCGALAYSSAAASISGQTGLVPMNNKGEG--GGALGFEFESSITONS	244
QY	239	YVLFENALTLTAIKGGVAVCCLPYSGSSTPPIYTFSPDNKOLVERHNSIMGALYAR	298
Db	245	SLFSEGNIA--IDAAGKGAIFYEKRG----EPIYTLISGNKSLTFAENSVYQGAICAH	299
QY	299	KLTISSGGPTLEFINNI SYANSONLGAIAIDTGETISLSAEKGTTFPOG--RTSLP--	353
Db	300	GLDLSAAMPTEFLSNRRCCNTAAGKGAIAIDSGSLSANOGDITFLGNTLTSTSAPTS	359
QY	354	FLNGHLIIONKFKLLOKRNCGSIEFYDPISEADGSGNO--LNIINDDPNKK--EYGTILF	410
Db	360	TRNATLYLSSAKITNLRKAAOGOSTIFPIASNTTGGASVOLTINOPDPSNPDIYSGITLF	419
QY	411	SGEKS LANDPR--DEKSTIPQVNL SAGYIYIKEGAETVYKFTQSPGSHLVLDLGTCL	467
Db	420	SGEKLSAEPKAKADPFTSILKOPALASGTLALKCNVELVDVNGFOTGEGSTILMOPGTCL	479
QY	468	IASKEDIAITGLAIDIDSLSSSTAAVIKANTANQOISVTDISIELISPGNAYEOLBRHN	527
Db	480	KADTEAISLTLYLDLSLEGKNSYSIETAG--ANKTITLTSLEPYQDSSGNFESHHTINO	538
QY	528	SOTPEPLSLERB--AGGSVTVYTAGDLPV--SPHAYFOGNMKLAFTGTGN--KVGEFFYDK	582
Db	539	AFTQPLVYFTATAASDIYIDALLTSPVOTPEPHGYOGHWEATWADSTAKSGTMTWYT	598
QY	563	INVKRPEKEGLVYNIILMGNAVNRSLMOQETHASSLQDRCGLWIDGIGNFEHVSASE	642
Db	569	TGYNPNPFRASVYFDSLMASFTDIRTLOQITSOANSTIYQOGRGLMASGTANFEHFKDSG	658
QY	643	DNIRKRNHSGGYLTVSNNEITPKHTYSMASFOLSEFRDQDAVANSNEYRYLGSYLYQYTT	702
Db	659	TNOAFRHKSYGYVGSAAEDFSENFISVAFQCLFGKDXDLFIVERTSHNYLASYLORRA	718
QY	703	SLGNIFRASKRPNPNVNVILSRFLQNLMLTFHLCAGHATNDMKTKTYANPMPKNSMR	762
Db	719	FLGGI-----PMSFESITDMLKDIPL--LINAOLSYTKNNDIDRTKSTPEAOGSNT	770
QY	763	NNCWAIECGSGMPLVFPENGRLLFOGALFPMKLOLYAYVAGQDFKETTADGRFRESNGLTSI	822
Db	771	NNSGALELGGSLATLYLPREAPFPQGYFPFLKQAYVNSQONKKEGAEARAPDDGDLVNC	830
QY	823	SVPLGIREKELASODVLYDFSFTIDPIFRKDPCECAALYISGDSMLVPAHVSRAHAY	882
Db	831	SIPVIRLEKISEEKEKNFEISLAIYIGVYRKNRPSRTSLWVSGASWTSILCKNLARQAF	890
QY	883	GSYGRYRHFNDYTELLCGSIECRPHANYNINCGSKFRF	922
Db	891	ASAGSHLTLSPHVELSGEAAEYLRKNSAITYNVDCLRKSF	930





[illegible]

RESULT 10  
B81591  
Polymorphic membrane protein G family, CP0308 [imported] - Chlamydophila pneumoniae (strain AR39)  
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence, revision 31-Mar-2000 #text, change 11-May-2000  
C:Accession: B81591  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: AB1500, MUID:20150255  
A:Accession: B81591  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-936 <REA>  
A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AAF38165.1; PID:g7189234  
A:Experimental source: Strain AR39, HL cells  
C:Genetics:  
C:Gene: CP0308

	Query Match	29.1%;	Score 1400.5;	DB 2;	Length 936;
	Best Local Similarity	35.9%;	Pred. No. 1,3e-76;		
	Matches 338;	Conservative 166;	Mismatches 391;	Indels 47;	Gaps 21;
Oy	PLVSLFLSLFSDLSLATTISLPEDSFHGDSONAERSYVQAQDVYLSTGVDYSINVD	68			
	: : : : : : : : : : : : : : : :				
Dd	PLFSLSLSIVA-EYTLDDSSNNSY---DQSNQTFTTVBSTTAAAGTTYSLSLDVSFQAAG	69			
Oy	--NSALNKACFNVTSGSVTFAGNHHGLYFNNISGGTKEGAVILCCDDQATARESGFSTL	126			
	: : : : : :     :				
Dd	ALGIPLASGCLEAGDLDTFOGNHAKLFAPINAGGS-AGTVASTSADRKMLPFNDFEURL	128			
Oy	SFIQSPDGIRE---OGCLYSKNAIIMLNINYREQNGSKKGAIISANTITYIGNDV	183			
Dd	SIISCPILLSPTQCAKLSYGNLSLTGNSOIIITQNSSNDGVYINKNLVLGSTSOFA	188			
Oy	184 SFYONAA---TFGGAIHSSGPLOIAVNOAEIRFAONTAKNGSGALXSDGIDIDIQNAVY	240			
	: : : : : : : : : : : :				
Dd	SFSNRQAFTKQGQGVAVATGTITTIENSFGIYSFSONLAK-GSGGAIVSITDNCSTIDNQY	247			
Oy	LFRNEALTTAIGKGAVACCLPTSGSSTPVPYIVFESDNKOLVFERHNISIMGGAIVARKL	300			
	:                 : : : : : : : : : : : : : : : :				

```

Dh 248 IFD$N$A$E$A$Q$G$G$A$IC$T$T$T$O$K$T$-----V$T$G$K$N$J$F$N$N$A$T$V$Y$G$A$IS$G$K$V 301
Qy 301 S$S$G$P$P$L$F$N$N$S$T$Y$A$N$-O$N$G$A$I$D$P$G$E$I$S$A$E$G$T$T$F$O$G$N$R$T$---S$L$P$L$N 356
Dh 302 S$S$A$G$P$P$L$F$O$S$N$S$G$S$A$G$G$G$A$N$A$S$A$E$A$T$A$S$T$G$D$T$F$N$N$O$V$T$N$S$T$R$N 361
Qy 357 G$H$L$O$N$K$F$K$L$A$O$R$N$C$S$I$E$F$D$P$T$S$--E$A$D$S$T$O$N$T$N$G$P$R$N$K$--E$Y$T$G$I$E$P$G 412
Dh 362 A$N$I$N$D$A$K$V$S$I$R$A$T$O$Q$S$I$Y$F$D$P$T$N$P$G$T$A$S$T$D$L$N$L$A$D$A$N$E$I$E$Y$G$A$I$Y$S$G 421
Qy 413 E$K$S$I$A$N$D$--P$D$E$K$T$T$P$O$N$V$N$S$A$G$Y$V$I$K$E$A$E$V$T$S$K$F$O$S$P$G$H$V$L$D$G$T$K$I$A 469
Dh 422 E$K$L$S$T$E$K$A$I$A$N$A$N$T$T$R$O$P$A$V$A$R$G$D$V$A$R$D$G$T$V$T$F$K$D$T$O$S$P$G$S$R$I$M$D$G$T$T$L$S$A 481
Qy 470 S$K$E$D$I$A$I$G$A$I$D$I$D$S$S$S$T$A$Y$A$N$K$A$N$K$O$S$V$T$D$S$T$E$I$S$P$G$N$A$Y$E$D$L$R$M$R$S$Q 529
Dh 482 K$E$A$N$S$T$I$N$G$L$A$V$N$S$S$D$G$T$N$K$A$-L$K$E$A$D$K$N$S$I$S$G$T$A$L$I$D$T$B$S$F$E$N$H$N$K$S$A$S 540
Qy 530 T$E$P$L$I$S$T$-E$P$A$G$S$V$T$Y$A$D$F$L$P$V$--S$P$H$Y$G$O$G$M$K$A$W$-T$G$T$N$K$Y$G$E$F$F$M$D$K$IN 584
Dh 541 T$Y$P$L$E$L$T$A$A$N$G$T$T$I$G$A$L$S$T$L$O$E$P$H$Y$G$O$G$M$O$S$M$A$N$A$S$K$G$S$N$W$T$P$T$G 600
Qy 585 Y$K$P$R$E$K$G$N$V$P$N$I$G$M$A$V$N$R$S$I$M$O$V$O$E$T$H$A$S$S$T$O$D$G$L$M$I$D$G$N$F$F$H$Y$S$A$S$D$N 644
Dh 601 Y$I$P$S$E$R$K$S$N$P$L$M$S$I$M$O$N$F$I$D$I$R$S$I$N$O$L$I$E$T$K$S$G$E$F$E$R$E$L$M$S$G$A$N$F$Y$R$D$S$M$P$T$R 660
Qy 645 I$R$Y$R$N$S$G$Y$V$L$A$V$N$N$E$I$T$P$K$H$Y$S$M$A$S$O$F$E$R$K$D$A$V$A$N$N$E$Y$B$M$Y$G$Y$T$O$Y$T$S$L 704
Dh 661 H$G$F$R$I$S$G$G$Y$A$L$G$I$A$T$T$P$A$E$D$Q$L$T$F$A$C$Q$L$F$A$R$D$R$N$H$T$T$K$N$H$G$D$Y$G$A$S$Y$F$H$H$T$E$G$L 720
Qy 705 G$N$I$F$R$Y$---A$S$R$N$P$N$V$G$I$S$R$R$E$L$O$N$P$L$M$I$P$H$F$L$C$A$Y$H$A$T$N$D$M$K$T$D$Y$A$N$P$M$V$K$N$S 760
Dh 721 F$D$I$A$N$F$L$W$K$A$T$R$A$P$---W$V$L$S$E$I$Q$I$P$L$S$-F$D$A$K$S$Y$L$H$T$N$H$M$K$Y$T$T$D$M$S$I$I$K$G$S 775
Qy 761 W$R$N$O$M$A$T$E$C$G$S$M$P$L$V$F$E$N$G$R$L$F$O$G$A$I$P$M$K$I$O$L$V$A$Y$O$G$D$K$E$T$T$A$D$O$R$R$S$N$S$G$T$ 820
Dh 776 W$R$N$D$F$C$A$D$A$S$I$D$-F$I$S$V$Y$P$L$K$E$Y$E$P$V$K$O$Y$I$A$H$O$O$D$E$Y$E$R$A$B$E$G$R$A$N$K$S$E$L$I 834
Qy 821 S$I$S$V$P$L$G$R$E$F$K$L$A$Q$D$V$L$D$F$S$F$Y$P$D$F$E$K$D$P$C$E$A$L$Y$S$C$D$M$L$P$A$H$V$S$K$H$A 880
Dh 835 N$V$E$I$P$D$V$T$E$F$R$D$S$E$K$G$Y$D$L$T$M$Y$I$L$D$A$Y$R$R$P$K$Q$T$S$I$A$S$D$A$N$M$A$Y$G$N$I$A$R$O$G 894
Qy 881 F$V$G$S$T$G$R$I$H$F$N$D$Y$T$E$L$L$C$R$G$S$I$E$C$R$P$H$A$N$Y$N$N$C$S$K$E$F$E 922
Dh 895 F$S$V$R$A$N$H$F$O$V$N$P$H$E$I$F$G$C$A$F$E$V$R$S$S$R$Y$N$N$T$G$K$E$C$F$E 936

```

RESULT 11  
C72078

polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CML025)  
C.Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C.Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C.Accession: C72078  
R.Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,  
Nature Genet. 21, 385-389, 1999  
A.Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A.Reference number: A72000; MUID:99206606  
A.Accession: C72078  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-936 <ANN>  
A.Cross-references: GB:AE001627; GB:AE001363; NID:g4376721; PIDN:AD18589.1; PID:g437  
A.Experimental source: strain CML029  
C.Genetics:  
A.Gene: pmp\_7

Query Match                      29.1%; Score 1400.5; DB 2; Length 936;  
Best Local Similarity    35.9%; Pred. No. 1.3e-76;  
Matches    338; Conservative    166; Mismatches    391; Indels     47; Gaps       21;

9 PLVSLTLLSVDFMTLSATTTISLRPEDSFGDSCONAKERSYVWQAGDYISLTGDVISINVD 68  
|||::: : : : : : : : : : : : : : : ||| |



```

Db 14 PLFSSLSIYAA-EVTLIDSSNSY--DGNGTFTVFTTDAAGTYYSLSDVSPQMAC 69
QY 69 --NSALNKACFNVTSGVTPAGNHGLEYFNINISSGTTKEGAVLCCODPQATARFSGFSTL 126
Db 70 ALGIPLASGCFLEAGDGLTQGNHAKFAFINAGSS-ACTVASTADKNLLEFNDSRL 128
QY 127 SFTQSPGDIKE---QGCLYSKNALMLNNYVRFEDONQSTKKGALISGANVTIVGNDYV 183
Db 129 SIISCPSLILSPGQCALKSVGNLSLTGNSQILFTQNFSSDNGGVINTRKNEFLSGTSQFA 188
QY 184 SFYQNA---TEGAIHSSGPLOIAVNOAETIRFAQNTAKKSGSGALYSDDDIDIDONAYV 240
Db 189 SFRNNAFTGKGGVYATYITTIENSPIVSPQNLAK-GSGGALYSTONCITTFNOY 247
QY 241 LFEENEALTTAIGKGVAVCCLPITSGSTPVPIVTFSDNKOJVEFRNHSINGGAIVARKL 300
Db 248 IFGNNSMWAQAQAGAICTTTDTK-----VTLGNKMLSTNNALTYGGAISLAKV 301
QY 301 SISSGPTLFINNIYSYANS-ONLGAIAIDTGEISLSAEKGTITFGNRT---SLPFLN 356
Db 302 SISAGPTLFGSNISGSSAGGGAIAINISAGELALSATSGDITFNNNGVTNGSTSTRN 361
QY 357 GHILONAKFLKQARNKCSIEFYDPTTS--EADGSTQNLNIDPNKK--EYTGTLIFSG 412
Db 362 ARIIDTAKVTSIRAAATGOSIYFYDPTTNGTAASDTLMLNADANSEIEYGGAIVFSG 421
QY 413 EKSLAND---PRDFKSTIIPONVNLISAGYLVIRKGAETVSKFTQSPGSHLYLGLFKLIA 469
Db 422 EKLSPEKAIANVTSTIROPAVLARGDVLKRGVYTFEDLTQSPESRLIMGGCTLLSA 481
QY 470 SKEDIAITGLAIDISLSSSTAIVIKANTANKOISVTDSEIELISPTGNAYEDLRMNSQ 539
Db 482 KEANLSINGLAIVNLSSIDGTNKAAL-LETEADKNISLSGTIALIDEGSFYENHNLSKAS 540
QY 530 TPELLSL-EEGAGGSVTVTAGDPLV---SPHYGQGNMKLAW-TGGNKVGEFEFMDKIN 584
Db 541 TYPLLEITTAGANGTITLGLSLTLTQEPETHYGOGNMGLSWANATSSKIGSIMWTRTG 600
QY 585 YKRPKEGNLVNINILMGNAVNRSLMOVOETHASSIQTDRGLMIDIGNEFFVASEDN 644
Db 601 YISPERKSNLPLNSLMGNIDIRSNOLLETSSGEPFERELMISIANFFFRDSMPTR 660
QY 645 IRYRHNSGVLSVNNETPKHAYTSMASFQSLFSRDXDAVANSNEYMYLGSYLQYTTSL 704
Db 661 HGRHRHSIGVALITGATTTPEDQLTFAFCOLFARDRNHITGKNHGDYTGASLFFHHTGL 720
QY 705 GNIFRY---ASNPNVNVGILSRFLQNLMLFHLCAIGHATNMKTYANFRPWKNS 760
Db 721 FDIANFLMKATRAP---WVLSISQIILPLS-FDAKFSTLHTDNHKKTYTTONSIIKGS 775
QY 761 WRNNAICAECCGSMPLLVFENGRLFOGAIIPMKQLQVYAYQGFKEETADGRRFSNGLT 820
Db 776 WRNDACADGADGLP-FVISVPLLKEVEFFVAVOYIYAQODFYERHAGRAFNKSELI 834
QY 821 SISVPLGIREFKIALISQDVLVDFESFYIPDIFRDKDSCAALVIGDSMLVPAHWSRHA 880
Db 835 INEIPIGVIFERDSKSEKGYDLTLMYIILDAYRRNPKCOTSLIASANMMAVGTNLA ROG 894
QY 881 FVSGSGTGRHENDYTELLRCGSTECRPHANVNLINGSKRFR 922
Db 895 FSVRAANHFOVNPHEIFGQFAFEVRSSSRNNTNLGSKFCF 936

```

RESULT 12  
 G86546  
 polymorphic outer membrane protein G family [imported] - Chlamydia pneumoniae (strain  
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
 C:Accession: G86546  
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A:Reference number: A86491; M01D:20330349

```

A:Accession: G86546
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-928 <STO>
A:Cross-references: GB:BA000008; NID:98978821; PIDN:BA08657.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: pmp_10

Query Match      28.9%; Score 1393; DB 2; Length 928;
Best Local Similarity 36.1%; Pred. No. 3.5e-76;
Matches 344; Conservative 162; Mismatches 374; Indels 74; Gaps 27;

QY 10 LVFSLTL--LSVDTLSATTTISLTFEDSPHGSQNAERS-YVNOAGDVYSLTGDSISN 66
Db 8 LVLSLTLCTSCSTYPAATAENIGPSDSFGSTNGTYTPKNTTGGIDYTLTGIDITLON 67
QY 67 V-DNSALNKACFNVTSGVTPAGNHGLEYFNINISSGTTKEGAVLCCODPQATARFSGFST 125
Db 68 LQDSALTKCFSDTTESSLFACKYSLSFLNLS--SAGGALSVYTDK-NLSLNGFSS 124
QY 126 LSTQSPGDI---KEQCLYSKNALMLNNYVRFEDONQSTKKGALISGANVTIVGND 181
Db 125 LTFELAAFSVITTPSGGAVKCGDILTFDNNGTILFKODCEBNGAISYKNSLKNSTG 184
QY 182 SVSFYQNAATE---GGAIHSSGPLOIAVNOAETIRFAQNTAKKSGSGALYSDDDIDIDON 237
Db 185 SIFEGKSSATKKGALCATGTVDITNTAPTLSSNNIAE-AAGCAIYSTONCITTFGN 243
QY 238 AYVLFRENEALTTAIGKGVAVCCLPITSGSTPVPIVTFSDNKOJVEFRNHSINGGAIVA 297
Db 244 TSLVFSNSVTATA-GNGCAL-----SGDAD---VTISGNQSVTFSGNOAVANGAIVA 293
QY 298 RKLSSIS--GGPTLFINNIYSYANSQNLGAIADTGEISLSAEKGTITFGNRT--TSLP 353
Db 294 KKLTLASGGGGGGSFNSNIVQGTAGNGAISILAAGECSLSAEADITFNGAIVATVP 353
QY 354 FL---NGIHLQNAKFLKQARNKCSIEFYDPTTSE--ADGSTQNLIN-GDPNKK-EYTG 406
Db 354 QTKRNSIDIGSKATKTNLRAISGHISFFPDITANTADSTLNLNKAADNNSYDSC 413
QY 407 TILFSGEKSLANDPR---DEKSTIIPONVNLISAGYLVIRKGAETVSKFTQSPGSHLYL 463
Db 414 SIYFSEKLSSEDAKADANLTLSTLKOPVTLTLAGNVLKRGVTLDTGFGTQFAGSSYIMDA 473
QY 464 GTKLISKEDIAITGLAIDISLSSSTAIVIKANTANKOISVTDSEIELISPTGNAYEDL 523
Db 474 GTTLKASTEEVTLTGSLIPVDSLGEER-KVYIAASAASKRVNALSPTLLIDNOGNAYENH 532
QY 524 RMNSQTFPLLSLEPGAGSVTVTAGDPLP---VSPHYGQGNMKLAW-----TGGNKV 575
Db 533 DLQKQDSFYVOL--SALGATITTDVPAYVTAPTHHYGQGWGWTWDDTASTRTKTK 580
QY 576 GEEFWDKINKYKRPKEGNLVNINILMGNAVNRSLMOVOETHASSIQTDRGLMIDIGNE 635
Db 591 ATLAWNTGTYLPBEROGPLVPSNLGSEFSDIOAIGVIERSLATLCSDFGFAAIVANF 650
QY 636 FHSASBEDNIRYRHNSGGVYLSVNNETPKHAYTSMASFQSLFSRDXDAVANSNEYMYL 695
Db 651 LDKDKGKKRYKHKGGGAIAGAAQTCSENLSIFPQGLFGSKDFLVAKNHTDTYAGA 710
QY 696 YLYQYTTSLGNIPIRYASRNPNVNVGILSR---FLQNPMLIFHLCAIGHATNMKTYD 751
Db 711 FYIQHTTEGSGF-----IGCLDLKLPESWGHKPLVLEGOL-AVSHVANDLTKTK 758
QY 752 ANFRPWKNSKRNNCAIECG---SMPLVFENGRLFOGAIIPMKQLQVYAYQGFKEET 808
Db 759 TAYPEYKSGSNNAFMMGLGASSHSYP---EYLHCFDYAPAPYIKNLTYIRDSSEK 814
QY 809 ADBRRRNSNGSLTISVPLGIREFKIALISQDVLVDFESFYIPDIFRDKDSCAALVIGDS 868
Db 815 TEGRSFDDSNLFWLSTPLGVKFEKFSDCNDSYDLTLSTVVDLRLNDPKCTALVIGAS 874

```



```

Db      346 DPEKRNAINIGSNGKFTTELRAAKNHTIEFYDPTISSEGTSSDVLIKINNGSAGALNPYOGTI 405
Oy      409 LFEGERSLANDPR---DEKSTIPONVNLASGYVVIKEGAEVTVSKFTQSGSHLVLDIGT 465
Db      406 LPSGETLTDELKVDADVDVNTDLSLIPVAKESKVI-VSGKLNLDIDEGNIYESHMF 465
Oy      466 KLIAKEDIATITGLAIDIDSLSSSTAAYIKANTANKQISVTDSELISPTGNAYEDLRM 525
Db      466 TLSTAGSTITITMLGINVDSLGLKOPVSLAKAGSNKVI-VSGKLNLDIDEGNIYESHMF 524
Oy      526 RNSQTPPLSLIEFGAGGSVTVTAGDELVPV-----SPHYGQGMWKLAM---TGTGKNVGE 577
Db      525 SHQUTSLKLTITVDADVNTDVLSSILIPVAPEDPNSBEYFGQGMVNNWTTDTATNTKEAT 584
Oy      578 FFWDKINXKPRPEKEGNLYPNILMGNAVNRSLMOVETIASSLQTDRLGIMIGINFEH 637
Db      585 ATWTXGTGVFVSPERKALVONTLMGVFTDIRSLQQLVEIGATGMEHKOGVWVSMNFIHL 644
Oy      638 VSASEDNIRYRHNHSGGYVLSVNNEITPK-HYTSMAFSQLEFSRDKDYAVSNEXRYMVLGSY 696
Db      645 KTGDENRKGFRHTSGGYVIG-GSAHTPKDGLFTFAFCHLPARDKCFIAHNSRTYGGTL 703
Oy      697 LVOYTTSLGINIFRYASRNPNVNGILSRFLQNP-----MIFHFLCAYGHATNDK 748
Db      704 FFKHSHTL-----QPQNYLRLGRAKFSESAIEKFPREIPLALDVOVSFHSNDRME 754
Oy      749 TDYANPVPYKNSRNRCMAIECGSMPLLVFENGRLFOGAIIPMKIQLVYAYOGDFKETT 808
Db      755 THYTSLEPESGWSNCEIAGIGIDLPEVLSNHPLEKFTIPQMKVEMYVSONSFESS 814
Oy      809 ADGRFRSNGSLTISVPLGIREFKLALSDVLYDFSFSYIPDIFFRKDPSCALVYISGDS 868
Db      815 SDRGRFSIGRLMLSLIPVAKFVQGDIGDSYTYDLGSEFVSDVYRNNPOSTATLVMSPPDS 874
Oy      869 WLVPAAHVSRHAFVSGTGRYHFNDDYTELLCRSICRPHARYNINCGSKFRF 922
Db      875 WKIRGNLSHQAFLLGRSNMYVNSNCELFGHYAMELRGSSRNMYNDVYKILRF 928

RESULT 15
D72077
polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: D72077
R:Kahn, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; M0ID:99206606
A:Accession: D72077
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-928 <ARN>
A:Cross-References: GB:AE001628; GB:AE001363; NID:q4376730; PIDN:AAD18593.1; PID:q437673
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: pmp_11

```

```

Query Match      27.9%; Score 1344; DB 2; Length 928;
Best Local Similarity 35.3%; Pred. No. 3.2e-73;
Matches 337; Conservative 152; Mismatches 385; Indels 80; Gaps 22;

```

```

Oy      10 LVFSLTLVFDLSATITSLTPEDSFHDSQNAERSYVWAG-----DYSITGD 61
Db      14 LAFSCHLQSLANDEL-----LSPDSSFNG-----NIDSGTTPKTSATYISLTGD 58
Oy      62 VSISNV-DNSALNKACFNVTSGSVTFAGNHGGLYFNNISGTTKEGAVLCCODPQATARF 120
Db      59 VFPEYEGKGTPLSDSCFKQTTDNLTFLGNCHSLTFEGIDAGTHAGAAASTTANKNLT--F 116
Oy      121 SGFSTLSFTQSPGD--IKRQGLYSKNALMLLNYYVREFQNSKTRGAGISGANVTYIVG 178

```

```

Db      117 SCFSLSDSPSPSTVYTGQGLSSAGVNLNIRLVANFNSTADGAIKASFLTLG 176
Oy      179 NYDSVSPYON-AATFGAIIHSSGPIQIAYVQAEIRFQONTAKNGSGALYSDGDIIDON 237
Db      177 TSGDALFNNSSSTKGCALITAGARIANTGYRFLSNAST-SGGALIDESTSILSN 235
Oy      238 AYLFRNEALTTAIGKGAVCCLPYSGSSTPVPYITFSDNKOVLERNHSGGAIYA 297
Db      236 KFLYFEGNAKTT-----GGAICWTKASGS-----PELISNNKTLIPASNAVETSGAIIHA 287
Oy      298 RKIISTSGGTLTI-NNISANSQNLGAIADTGEIISAKKGTITROGN-----RT 350
Db      288 KKLALSSGFTTEFLRNNVSATPK--GGAISIDASGSLSAETGATTFRNLTTTGST 345
Oy      351 SLPLINGIHLQNAKFLQIARNGCSIEFYDPTISSEADSGTQOLINGDPKN--KEYTGT 408
Db      346 DPEKRNAINIGSNGKFTTELRAAKNHTIEFYDPTISSEGTSSDVLIKINNGSAGALNPYOGTI 405
Oy      409 LFEGERSLANDPR---DEKSTIPONVNLASGYVVIKEGAEVTVSKFTQSGSHLVLDIGT 465
Db      406 LPSGETLTDELKVDADVDVNTDLSLIPVAKESKVI-VSGKLNLDIDEGNIYESHMF 465
Oy      466 KLIAKEDIATITGLAIDIDSLSSSTAAYIKANTANKQISVTDSELISPTGNAYEDLRM 525
Db      466 TLSTAGSTITITMLGINVDSLGLKOPVSLAKAGSNKVI-VSGKLNLDIDEGNIYESHMF 524
Oy      526 RNSQTPPLSLIEFGAGGSVTVTAGDELVPV-----SPHYGQGMWKLAM---TGTGKNVGE 577
Db      525 SHQUTSLKLTITVDADVNTDVLSSILIPVAPEDPNSBEYFGQGMVNNWTTDTATNTKEAT 584
Oy      578 FFWDKINXKPRPEKEGNLYPNILMGNAVNRSLMOVETIASSLQTDRLGIMIGINFEH 637
Db      585 ATWTXGTGVFVSPERKALVONTLMGVFTDIRSLQQLVEIGATGMEHKOGVWVSMNFIHL 644
Oy      638 VSASEDNIRYRHNHSGGYVLSVNNEITPK-HYTSMAFSQLEFSRDKDYAVSNEXRYMVLGSY 696
Db      645 KTGDENRKGFRHTSGGYVIG-GSAHTPKDGLFTFAFCHLPARDKCFIAHNSRTYGGTL 703
Oy      697 LVOYTTSLGINIFRYASRNPNVNGILSRFLQNP-----MIFHFLCAYGHATNDK 748
Db      704 FFKHSHTL-----QPQNYLRLGRAKFSESAIEKFPREIPLALDVOVSFHSNDRME 754
Oy      749 TDYANPVPYKNSRNRCMAIECGSMPLLVFENGRLFOGAIIPMKIQLVYAYOGDFKETT 808
Db      755 THYTSLEPESGWSNCEIAGIGIDLPEVLSNHPLEKFTIPQMKVEMYVSONSFESS 814
Oy      809 ADGRFRSNGSLTISVPLGIREFKLALSDVLYDFSFSYIPDIFFRKDPSCALVYISGDS 868
Db      815 SDRGRFSIGRLMLSLIPVAKFVQGDIGDSYTYDLGSEFVSDVYRNNPOSTATLVMSPPDS 874
Oy      869 WLVPAAHVSRHAFVSGTGRYHFNDDYTELLCRSICRPHARYNINCGSKFRF 922
Db      875 WKIRGNLSHQAFLLGRSNMYVNSNCELFGHYAMELRGSSRNMYNDVYKILRF 928

```

Search completed: October 2, 2001, 03:29:00  
Job time: 1317 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2001, 03:32:05 ; Search time 48.09 Seconds  
(without alignments)  
656.759 Million cell updates/sec

Title: US-09-446-677B-6

Perfect score: 4814  
Sequence: 1 MRFSLCGPPLVFLSLVSVF.....IECRPHANNYNINGSKRFR 922

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	567	11.8	964	1	OMPE_CHLTR
2	393	8.2	1034	1	OMPE_CHLTR
3	202.5	4.2	1250	1	YFAL_ECOLI
4	198.5	4.1	1286	1	ALDA_ECOLI
5	196	4.1	2249	1	190K_RICRI
6	177.5	3.7	1643	1	OMPE_RICPR
7	175	3.6	1656	1	OMPE_RICPA
8	173.5	3.6	1654	1	120K_RICRI
9	173.5	3.6	1654	1	OMPE_RICRI
10	172	3.6	1645	1	OMPE_RICRY
11	170	3.5	2003	1	YDBA_ECOLI
12	168.5	3.5	1569	1	YFPA_ECOLI
13	168	3.5	1577	1	HLVA_PROMI
14	164.5	3.4	1608	1	HLVA_SERMA
15	161.5	3.4	995	1	YICQ_YEAST
16	157.5	3.3	1848	1	CBPA_CLOCL
17	156	3.2	2334	1	MAPA_BACSU
18	155	3.2	1140	1	YMG6_YEAST
19	151.5	3.1	3591	1	FHAB_BORPE
20	149	3.1	1325	1	YDER_ECOLI
21	148	3.1	1394	1	HAP_HAEIN
22	147	3.1	1025	1	SLAP_CAUCR
23	146.5	3.0	1743	1	TAGC_DICDI
24	145	3.0	888	1	LACG_DICDI
25	145	3.0	1545	1	IGAI_HAEIN
26	142	2.9	851	1	BCK2_YEAST
27	141	2.9	881	1	YEER_HAEIN
28	140.5	2.9	1122	1	ADP1_MYCGA
29	138.5	2.9	1541	1	IGAI_HAEIN
30	138	2.9	1039	1	AG43_ECOLI
31	136	2.8	1306	1	MSB2_YEAST
32	135	2.8	952	1	UVRA_MYCGE
33	135	2.8	1034	1	ICEN_PANAN

34	134.5	2.8	781	1	NANH_VIBCH
35	134.5	2.8	910	1	PERT_BORPE
36	134	2.8	917	1	HXA3_HAEIN
37	133.5	2.8	934	1	Y321_MYCGE
38	133.5	2.8	1567	1	ICEN_XANCT
39	133	2.8	1041	1	EGT2_YEAST
40	133	2.8	1332	1	ICEN_PANAN
41	131	2.7	1183	1	YMG6_YEAST
42	130.5	2.7	821	1	YICQ_ECOLI
43	130.5	2.7	854	1	VG12_BPPH2
44	130	2.7	1061	1	OAR_MYXAX
45	130	2.7	1228	1	SLAP_BACST

## ALIGNMENTS

RESULT	ID	OMPE_CHLTR	STANDARD	PRT	964 AA.
AC	084877	OMPE_CHLTR			
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DE	PUTATIVE OUTER MEMBRANE PROTEIN E PRECURSOR.				
GN	PMPE OR CT869.				
OS	Chlamydia trachomatis.				
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.				
OX	NCBI_Taxid=813;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=D/UM-3/CX;				
RX	MEDLINE=99000809; PubMed=9784136;				
RA	Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,				
RA	Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koorn E.V.,				
RA	Davis R.W.;				
RT	"Genome sequence of an obligate intracellular pathogen of humans:				
RT	Chlamydia trachomatis."				
RL	Science 282:754-759(1998).				
CC	- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)				
CC	(POTENTIAL).				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
DR	EMBL: AEO01360; AAC68467.1; -				
KW	Outer membrane; Signal.				
FT	SIGNAL	1	18	POTENTIAL.	
FT	CHAIN	19	964	PUTATIVE OUTER MEMBRANE PROTEIN E.	
SO	SEQUENCE	964 AA; 104703 MW; 1B998A7D2E57ACE2 CRC64;			

Query Match 11.8%; Score 567; DB 1; Length 964;

Best Local Similarity 24.6%; Pred. No. 7.7e-27;  
Matches 236; Conservative 147; Mismatches 358; Indels 220; Gaps 43;

QY	57	SLTGDV-SISN--VDN-----SALNK-----ACFNVSQSVTFGNHGLYFNINISGTT 103	
DB	45	SLTGDVHNLNCTYDNLRYLALITQKTPNMGAAVTTDYLSFPTQKEGILF--ANKLT 101	
QY	104	KE--GAV-----LCC-----ODPATARFSGFSLTSFQSP 132	
DB	102	PESGAGCYVSPNSPYEIRDTIGVIFENNTCCRLTWTNPPYAADK----- 148	
QY	133	GDIREGCLYSKNAKMLNNYVRFEDNOSKTKGAGISANVTIVGN-----YDSVSF 185	
DB	149	--IREGAIHAONLYINHNDVYVGFMAKFSYVOGAI StantonFVSNOSCFLEMDNICI 206	

```

QY 186 YQNAATFGAIIHSSGPIQIAVNOAEIRFAONTAKNSGALYSO-----GDIDIDON 237
QY 207 QYTWAGGAGIAYAGTSNFSFNNDLFFINNAC--CAGAIFFSPICSLTGNRGIVFYNN 264
QY 238 AYVLFRENEALTAIGGAV-----CCLPTSGSSPPVIVFSDONKOLVFERNHSIMCGA 294
QY 265 R--CFKNVETASSASDGAIKVTTRLDVTGNGR---IFESDN---ITKNY---GGA 311
QY 295 IYARKLSSISSGGPTLFINNITSYANSONIQAIAIDTGEISLSAKETITF----- 345
QY 312 IYAPVYLVNGPTLYFINNI---ANKKGAIYITDTSNKSISDRALIFENIYIVNT 367
QY 346 --OGNRTSL--PLINGIHLIQAIAKFLQAIRNCGSIEFYDPI--TSEADSGSTQIUNGDPK 400
QY 368 NANGSTSNAPPRNATITVASSGEIILGAGSSONLIFFIPIEVSNGAVSVFN-----K 422
QY 401 NKREYTGILISGEKSLAND--PRDFKSTIPONVNLGAGYLIVKEGAVVYSKPTQSGSH 456
QY 423 EADQTSVSVSGATVNSADHQRMLQKTPAPLTLISNGFICIEHQADLVNRFOTG-- 480
QY 459 LVLDIGTKL-----ASKEDIAITGLAIDISLSSSTAA---VIKANTANK- 502
QY 481 -VVSIGAGVLSCKYKNGTGSASASATLTKHIGLNLISLKSAGELPLVWEPTNNNNY 539
QY 503 -----QISVTD--SIELISPTGNA--YEDLRMNS--QTEPLISLEPGAGSVTVYAGDPL 553
QY 540 TADTAATFSLSDVKLSDIDYGNSPYESTDITFHALSSQPMLSISEASDNOLOSENIDFSG 599
QY 554 VS--PHYGFOGNWKLAMTGTGN-----KVGEFF-----WDKINTYKPREKEG 593
QY 600 LANPHIGWOLKWTGNAKTDDEPPASATITTDOKANRFRITLLTYLPAGIYVSPKHS 659
QY 594 NIVPNILMGAVNVRSIMQVETHASSIQTRGLMIDIGNFHVSASEDNIRYRHSNG 653
QY 660 PLANTLMGMMLATESLKSASALTEPSGHPFWGITGGGLAMNYQDPREHMPGHMSSG 719
QY 654 YVLSVNNETPKYITSAEQLSR--DKDYAVSNNEYRMT--IGSYLYQ-----YITSLG 705
QY 720 Y--SAGMIACQHTFSLKFSQYTKLNERAKNNVSKNSCOGEMFLSQEGFLLTKLV 777
QY 706 NIFRYSARNPNVNVGLSRFLQNPIMIFHFLCAYGHATDMKTDVANFPMVKNMNNC 765
QY 778 GLXSYGDH-----CHHFTOGENLTSOCTF----- 804
QY 766 MAIECGSS--MPLVFENGRLFOGAIPEMKLQIYVAYOGDEKETTADGRFSNGS--LT 820
QY 805 -SQTMGAVFEDLPMPKPFSTHIL--TAPFLGALGITYSSLSHTEVGAYRFSSTKTPLI 861
QY 821 SIVPLGIREKL--ALSQDVLDPFSFSYIPDIRKDPSCAALVIGSDSLVPAAHVSRH 879
QY 862 NVLVPITGVKGSFNATHRPQAMTVETALYQVLYROEPGIAAQLASKGIWFGSGSPSSRH 921
QY 880 A 880
QY 922 A 922

```

```

RESULT 2
OMPF.CHLTR STANDARD: PRT: 1034 AA.
AC P38008; 084878;
DT 01-OCT-1994 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN F PRECURSOR.
GN PMPF OR CPM870.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=D/TW-3/CX;
MEDLINE=9000809; PubMed=9784136;

```

```

RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
RN [2]
RP SEQUENCE OF 26-35.
RC STRAIN=L2/434/BU;
RA Bini L., Santucci A., Magi B., Marocchi B., Sanchez-Campillo M.,
RA Comanducci M., Christensen G., Birkelund S., Vitreou E., Rattl G.,
RA Pallini V.;
RL Submitted (SEP-1994) to the SWISS-PROT data bank.
CC -SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AEO01360; AAC68468.1; -
KW Outer membrane; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1034 PUTATIVE OUTER MEMBRANE PROTEIN F.
FT CONFLICT 34 35 RR -> FH (IN REF. 2).
SQ SEQUENCE 1034 AA; 112392 MW; 445FF4C3D463AE7 CRC64;

```

```

Query Match 8.2%; Score 393; DB 1; Length 1034;
Best Local Similarity 21.1%; Pred. No. 2.9e-16;
Matches 231; Conservative 142; Mismatches 418; Indels 306; Gaps 45;

```

```

QY 13 SILLVSFDTSLATISLTLPEDSFHDSQNAERSYVQAGDYVSLTGDVISNVDNAL 72
QY 57 NIVLSNLQNSGTACTISGTOTQIFSNVNT-----TADGGAGFDMV--TTSFASDANL 111
QY 73 --NKACFN-----VTSQSVTFAGNHGLYFNINISSGTTEKGAVALCCDPQATAFSG 122
QY 112 LFCNNCTHKKGGAGAIRSGPIRFLNNDVLFYNNISAG-----AKYVG 155
QY 123 FSTLSFSTQSPDIKEQ--GCLYSKNAIMLLNNYVFEONOSKTKGAISG--ANVTIVG 178
QY 156 -----TGDHNEKRRGALYA--TTITLTGNRTIAFINMGGDCGAIADTQISITD 205
QY 179 NYDSVSFYQNA-----TFGGAIIHSSGPI--QIIVNOAEIRFAONTAKNSGGA 225
QY 206 TVKGILFENHNTLNHLIPYQAEENMARGAICSRRLCSISNNSGPIVFYFN--OGKGGGA 263
QY 226 IYSDGDIDIDQNAVYLFRENEAL--TTAIGKGAVCCLPYSGSSTPVYIVTFSDNK--Q 280
QY 264 ISATRCYIDNNKERIIFSNSSLSGSSOSSASNGAL--QTTG-----FLRNNKGS 314
QY 281 LVEERNHSIMGGAIAARKLSSISGGPTLFINN-----ISYANSONIQAIAIDTGEIS 335
QY 315 IYFDSMTATHAGALINGYIDIRDNGPVYFLNNSAAMGAFAFNLSKPRSATNTYHTG---- 370
QY 336 LSAEKGTITFGKRTSLPLINGIHLIQAIAKFLQAIRNCGSIEFYDPI 383
QY 371 -----TGDIVFNNN--VFETLDGNLGRKRLFIHNNNETPYTLISGAKKDRITRYEDLF 423
QY 384 TSEADSGSTQIUNGDPK-----EYTGILISGE-----KSLANDPRDFKSTIP 428
QY 424 QWERVKNENSNPPSPTSRITIVNPEETESGAVVSYNOMSSDITLMKKEINIKRPA 463
QY 429 QNVNLSAGYLIVKEGAEVYSK--FTQSPGSHVLVDGKRLASKK--DIAITGLAIDI-- 483
QY 484 --TLKFGTLAIDDALEIFENFTQNPSTLALSGATLVGKRGKLNITMGLYLP 541
QY 484 -----DSLSSSTAAVIKANTANKQISYVDSIELISF 515

```

```

Db 542 ILKRGSPCIRVNPQDMQNTGQTSPSSSTISPTMII-----FNGRLSIDV-----EN 592
QY 516 TGAAYEDLRNRNSTEPL-LSLEPGAGGCVTVAGDPLPS-----PHYGQGNMKLAW-- 568
Db 593 YESVYDSMDLSRGAEOQLISTETNDGOLDMSWQSSLANSLSPPHYGYOGVMTWMTIT 652
QY 569 -----TGNGKXGEEF 578
Db 653 TTYTITLNNNSAPTSTSTAEOKTSETTPSPNTTASTIPNKAASGSGSHNSGEV 712
QY 579 F-----WDKINRKPREKEGNLVPNIIANGNANV-----VNSIAMQ-----VOETHASS 620
Db 713 TITKHTLVWMAAPVGYIVDPIRGGDILANSLVSGRMWMTGLSLDPSNMPALQGAAT 772
QY 621 LOTDRGLMIDGICGNFPHVASSEDNIRYRNHNSGGYVLSVNNETPKHYTSAFQSLESRD 680
Db 773 LFTKQKRLSYHG-----YSSASKGYTVSSQASGAGHGFLLSFSSQSSDKM 819
QY 681 DYAVSNN-EYRMYLGSYLXYOTYSLGNIFRYASRNPNVNGILSRRLQNPIMFHLCA 739
Db 820 EKETNRLSRVYLSALCFEHP-----MFDKALIGAAACN 855
QY 740 YGATNDMKTQDYANFPVKNKSMRNMCALIEGSGMPLVFEENGRLFGAIPFKLQLVYA 799
Db 856 YG--THMRSFYG---TKRSKGRFHSHTLGLASLRCEL-----RDSMPLMSIMLTPF 902
QY 800 YGDFKETTADGRRFSGLS-----TSISVPLGIFEXKALASQD-----VLNDFSF 845
Db 903 AQAALFSTREPASIR-ESGDARLFTLEQAHHTAVASPIGI---KGAVSSDTPWPLSWMEL 958
QY 846 SYIPDFRKPSCSEALVVISGDSMLVPAHVSRAHAFVSGTGRHFNDYTELLCRGSI 905
Db 959 AYQPTLYMKRPLANTLILQNGNSWVTNTPLAHSPFGRSLSLKS-HLKLPAVNOAEV 1017
QY 906 RPHARNYNINCGRKFR 922
Db 1018 ATSTVSHYNAGALVF 1034

RESULT 3
YFAL_ECOLI
ID YFAL_ECOLI STANDARD; PRT: 1250 AA.
AC P45508: P45507; P45506; P39441; P76468; P77487;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 131.2 KDA PROTEIN IN UBIG-NRDA INTERGENIC REGION
DE PRECURSOR.
DE YFAL.
GN YFAL.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBL_taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Berna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K., Kasai H.,
RA Kashimoto K., Kim S., Kinura S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizouchi K., Mori H., Motomura K., Nakamura Y.,
RA Nasuboto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA Tegan H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

```

```

RN [3]
RP PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
RX MEDLINE-84272624; PubMed-6087316;
RA Carlson J., Fuchs J.A., Messing J.;
RT "Primary structure of the Escherichia coli ribonucleoside diphosphate
RT reductase operon."
RL Proc. Natl. Acad. Sci. U.S.A. 81:4294-4297(1984).
RN [4]
RP SEQUENCE OF 925-1198 FROM N.A.
RC STRAIN-K12 / EMG2;
RA Estep P., O'Keefe T., Robison K., Church G.M.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1180-1250 FROM N.A.
RC STRAIN-OV6;
RX MEDLINE-88201664; PubMed-2834621;
RA Hussain K., Elliott E.J., Salmond G.P.C.;
RT "The pard-mutant of Escherichia coli also carries a gyrA mutation.
RT The complete sequence of gyrA."
RL Mol. Microbiol. 1:259-273(1987).
RN [6]
RP IDENTIFICATION.
RX MEDLINE-96032851; PubMed-7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes."
RL Nucleic Acids Res. 23:3554-3562(1995).
CC -1- SIMILARITY: TO E.COLI YDEK.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
CC MANY FRAMESHIFTS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AE00313; AAC75293.1; -.
DR EMBL: D90855; BAA16052.1; ALT_INIT.
DR EMBL: D90854; BAA16050.1; ALT_INIT.
DR EMBL: K02672; -. NOT_ANNOTATED_CDS.
DR EMBL: U30459; AAA74094.1; -.
DR EMBL: Y00544; -. NOT_ANNOTATED_CDS.
DR EcoGene; EG12850; yfal.
KM Hypothetical protein: Repeat: Signal.
FT SIGNAL 1 23
FT CHAIN 24 1250
FT DOMAIN 919 948
FT FT 28 30
FT CONFLICT 40 40
FT FT 65 66
FT CONFLICT 431 431
FT FT 433 434
FT CONFLICT 478 478
FT FT 773 773
FT CONFLICT 853 853
FT FT 923 924
FT CONFLICT 948 994
SQ SEQUENCE 1250 AA; 131152 MW; 17F98C05E299FC95 CRC64;

```

```

Query Match 4.2%; Score 202.5; DB 1; Length 1250;
Best Local Similarity 23.7%; Pred. No. 0.00013;
Matches 130; Conservative 83; Mismatches 197; Indels 139; Gaps 30;
QY 100 SGTTKGAVLCODPOATANFSGFSTLQSPGDI-KEQGCLYS-----KNALMLINNY 153
Db 51 SGIQOWSI-----ADGQWLFSDMTNNAAGAVFLQGAFFSLIPENETMTLEFANN 103

```

```

OY 154 VVFEONOSKTKGAI--SGANVTIVGNVDSVSYON--AATFGGAHSS-----GPIQIAV 206
DB 104 IYTGELN-----NGALFAKENSTL--NLNDVITISGNVAGYGAITYSSGNDGAVDIR 157
OY 207 NOAEIFPQNTAKNSGGALYSDGIDIDONAY---VLFRENALPT--AIKKGAVCC 260
DB 158 TNA--MFRNNIANDGKGAIYT-----INNDVYLSVIFDNNQAYSTSYSDGDGAIDV 210
OY 261 LPTSSSTPIVITFSDNQVLVERNHSIMGGAITYARKISTISGGPTLEINNISYANSQ 320
DB 211 TDNNSSKHPGTYIVNN--TAFNTAEGYGAITYNSVTAP-----YLDISVDDST 262
OY 321 NLGGAIAID-----TGGEISLSAEKGTITPQGNRTSLPFLNGIHLQNAK 365
DB 263 SQNGVVLVDENNAGYAGDGPSSAGGFMYLGLSEVTFPIADKTLV-----IGNENNDG 317
OY 366 FLKQARNGCSIEFYPTISEADGSTOLNINCPKKEITG-----TILF 410
DB 318 AVDSIAGTGL-----ITK--TGSGLVLYNAD--NNDFTGEMQIENGVEYTLGRNSLNM 366
OY 411 SGEKSLANDPRD-----FKSTIPQNV-----NLSAGVLYIKEGAEVT 447
DB 367 VGTGHCODPDQCYGLTIGSIDOYQNAELNVSTQOTFVHALTGQNGTLNIDAGNVT 426
OY 448 VSK-----FTQSPGSHLVLDLGTKLIAKEDIAITG--LAIDIDS--LSSSTAIVIRANTA 500
DB 427 VNGSPAGIIEGAGQLTQNGSYVLAGQSMALTGDIYVDDGAVLSBEDADIALAD 486
OY 501 NKOISTYDS--ILEIS---PTGNAYED--LRMNSOTFPLLSLEPGAGSVYTAGDFLP 553
DB 487 DPOSIYLVNGVLDLSDFTWQSTGYNDGLEVSGSGTYI-----GSQDVVYDLAGGDNL 540
OY 554 VSPHYGEG 562
DB 541 ---HIGGDC 546

```

```

RESULT 4
AIDA.ECOLI
ID AIDA.ECOLI STANDARD; PRT: 1286 AA.
AC 003155;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE ADHESIN AIDA-I PRECURSOR.
GN AIDA-I.
OS Escherichia coli.
OC Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RC STRAIN=2787 (O126:H27);
RX MEDLINE=92326638; PubMed=1625582;
RA Benz I., Schmidt M.A.;
RT "AIDA-I, the adhesin involved in diffuse adherence of the
RT diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is
RT synthesized via a precursor molecule."
RL Mol. Microbiol. 6:1539-1546(1992).
RT
RT FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
RT ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
RT TO EPITHELIAL CELLS.
CC -I- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

```

DR EMBL: X65022; CAA46156.1; -.
DR PIR: S28634; S28634.
KW Cell adhesion; Signal; Outer membrane; Plasmid.
FT SIGNAL 1 49
FT CHAIN 50 7 ADHESIN AIDA-I.
FT PROPEP 2 1286
SQ SEQUENCE 1286 AA; 132271 MW; B2A00F72AC05FB34 CRC64;

```

```

Query Match 4.1%; Score 198.5; DB 1; Length 1286;
Best Local Similarity 19.3%; Pred. No. 0.00024;
Matches 198; Conservative 135; Mismatches 354; Indels 337; Gaps 51;

```

```

OY 46 RSYVOAGDVYSLTGVNLSN-----VDNSA-----LNKACFVNTSG 82
DB 352 RKGISAGTIVNOSGYVNISGGYASTIINSGLTLYSDGARGTILNNSGREVNSG 411
OY 83 SVTF-----AGNHGLY-----FNNISGTTK--EGAVLCCQDPQATAR-- 119
DB 412 GVSYNAMINTGNQYIYSDGEATAIYNTSGFORINSGTAPQVNSVYVTRVSSAKPF 471
OY 120 -----PSGFSTLSFOSPGDIDEGGLYSKMLMLNLYVFEON 160
DB 472 DAEVYSGKQTYVLMRGVWSNFLTAVMSMFPCTASGANVLSGRNAPAGNVVGTILNQ 531
OY 161 QSKTKGAGISGANVT--IVGNVDSVSEYQNAATFGAIIHSSGPIQ-----IAVNOA 209
DB 532 EGRQY--VYSGATATYTVGNNGREYVLSGGLTIDGYVLSNGGLQAVSSGKASATYINEG 589
OY 210 EIRF-----AONTAKNSGGALYSDGIDIDONAYVLFRENALTTAIGKGAVCCLP 265
DB 590 GAQFVYDGGVGTNTIKNG-----GITRVDSGA-----SALNIALSSGN--LFTST 634
OY 266 STTPVPIVTFSDMKOLVFEENHS---IMGGAITYARKLSSGGPFLINNISYANSQ 322
DB 635 GATLPETITWA--ALVSQNHASNIVLENGL--LRVTSGG-----TATDTVNS 680
OY 323 GGAIAIDTGEISLSAEKGTITPQGNRTSLPFLNGIHLQNAFLKQARN----- 373
DB 681 AGRLRIDDGITIN-----GTTINAD--GIYAGINIQDGNFIILNLENVFEETELSG 721
OY 374 -----GCSIERYDPTTSADG-----STQALING 397
DB 732 SGVLVNDNTGIMTYAGTLTQAGVNVKNGIIPDSAVNVANMAVNONAYINISDQATING 791
OY 398 DPKNKEYTGRIEFSG---EKSLAND-----PRDEKSTIPONVNLASGYLVIKES 443
DB 792 SVNN---NGSIVINNSIINGNTINDADLSFGTAKLLSATYNSGLVNNKNIILNPTRESAG 848
OY 444 AEVTSKFTQSPGSHLVLDLGTKLIAKEDIAITGLAIDIDSLSSTAIVIRANTANKO 503
DB 849 NLTLYSNYCTPGS--VYSLGVL---EGDNSLTDLRVYKNTSGSDIYVVEDSGGG 903
OY 504 ISVTDSIELISPTGNAYEDLRMNSOTFPLLSLEPGAGSVYTAGDFLPVSHYGFQ-G 562
DB 904 --TRDINIISVGNDAEFSLNK-----RVAAAY-----DYTLQNG 939
OY 563 NKLAMTGNKNKYGEEFWXINKPK--RPEKGNLVPIILMGNAVNVNSLQVOYE 615
DB 940 N---BSGTDNK---GWIYTLSHLPSTDRQYRPE--NGSIAIYMALANSFLMDLIERKO 990
OY 616 THASSLOT--DRGLWI-----DG-----IGNFE 636
DB 991 FRAMSDNTQPEASVMAKKTIGTSSGKINDGQKTTTNOFINLGDDIYKFAHAGQDGT 1050
OY 637 -----HVSASEDNIRYRHNSG-----GYVLSV-----NNEITPKHY--TSMASQLF 676
DB 1051 LGIMGYANAKKGTINTYNSKKAARNTLDGYSVGYCTWYONGENATGLAETWQYNNFN 1110
OY 677 SRPKDYAVSNNEKRMATLGSILYQTTSLGNIERYASRNRYVY-----GLTSRFEION 729
DB 1111 ASYKGDGLEEKYN-----LNGLTASAGGY-----NLNVHTWTSBPGITGEFMIAP 1157

```



QY 730 PLMIFHFLCAYGHATNDKTDYANFPMVKNSMHN-----CMAIEGSGMPLIVFE 780  
 Db 1158 -----HQAAMKGVPTPTHOE-DNGTVAGAGKNNIOTKAGIRASMKVK-----STLDKD 1206  
 QY 781 NGRIFGALFPMKQLVAYAGDEKETTADGRFNSGLTSISVPLGIR---FEKLALSQ 837  
 Db 1207 TGRFR---PYIANNMIHNTH-EFGVKMSDDSLSSRNGEIKNGIEVIONLSVNG 1262  
 QY 838 DVLY 841  
 Db 1263 GVAY 1266

RESULT 5  
 190K\_RICRI STANDARD; PRT; 2249 AA.  
 ID 190K\_RICRI  
 AC P15921;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).  
 OS Rickettsia rickettsii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_Taxid=783;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R;  
 RA MEDLINE=90354033; PubMed-2117568;  
 RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;  
 "A protective protein antigen of Rickettsia rickettsii has tandemly  
 repeated, near-identical sequences.";  
 RL Infect. Immun. 58:2760-2769(1990).  
 CC - FUNCTION: ELICITS PROTECTIVE IMMUNITY.  
 CC - PTM: GLYCOSYLATED (POSSIBLE).  
 CC - DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF  
 AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@sib-sib.ch).  
 CC -----  
 DR EMBL: M31227; AAA26380.1; -  
 DR Antigen; Repeat; Signal; S-layer; Glycoprotein.  
 KW PIR: A41477; A41477.  
 FT SIGNAL 1 28  
 FT CHAIN 29 2249 190 KDA ANTIGEN.  
 FT DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 212 286 A (TYPE I).  
 FT REPEAT 287 358 B (TYPE II).  
 FT REPEAT 359 430 C (TYPE II).  
 FT REPEAT 431 505 D (TYPE I).  
 FT REPEAT 506 577 E (TYPE II).  
 FT REPEAT 578 652 F (TYPE I).  
 FT REPEAT 653 724 G (TYPE II).  
 FT REPEAT 725 799 H (TYPE I).  
 FT REPEAT 800 874 I (TYPE I).  
 FT REPEAT 875 949 J (TYPE I).  
 FT REPEAT 950 1021 K (TYPE II).  
 FT REPEAT 1022 1093 L (TYPE II).  
 FT REPEAT 1094 1165 M (TYPE II).  
 FT REPEAT 1166 1180 TYPE I (INCOMPLETE).  
 SQ SEQUENCE 2249 AA; 224333 MW; A9D64646C089DF087 CRC64;

Query Match 4.1%; Score 196; DB 1; Length 2249;  
 Best Local Similarity 19.6%; Pred. No. 0.00075;  
 Matches 217; Conservative 111; Mismatches 369; Indels 410; Gaps 51;

QY 24 LSATTSIPEDSEFHDSQAERSYNOAGDYSLTGDSVLSNVDNSALNKACPNYSGS 83  
 Db 898 IKATYTKLTNAASV---LFLTNAAVLTAIDFTTGG---DNQVNLNGLALSQVLT-GD 949  
 QY 84 VTFAGNHGILYFNINSSGT-TKEGAVLCODPOTARFSGFTLSIOSGDIKEOGLY 142  
 Db 950 I---GNTNSLATISVAGCTLTGAVIKATTTKLTDAASAVKFTNPVYVGAIDNNG--- 1003  
 QY 143 SKNALMLNNYVVFEEONOST-----KGAIISGANVTYGVND 181  
 Db 1004 --NA---NNGIYTFGTGNSVTGNVGTNATLATVYVAGLLOYGVGVKANTINTDNAS 1057  
 QY 182 SVSEYONAAFFGAIHSSGSLQIAVNOAEIRPAONTAKNS----- 222  
 Db 1058 AVTF-TWVVVVTGAIDNTG---NANGIYTFGTGNSVTGNVGTNATLATVYVAGLLOY 1112  
 QY 223 -GGLYSYD-----GDIDIDONA---YVLFRENEALTTAIGKGV 258  
 Db 1113 QGGVAVANTINTLNDMSAVFTNPVYVGTALDNTGNANNGIYFTGNSVTGIDIGNAL 1172  
 QY 259 CCLPT-----SGSSTPVPDYTFPSDNKQLVFERNHSIMGG-AI-YARKLSISSGGPTL 309  
 Db 1173 ATYVVGAGITLQAGGSLAANNIDFGARSTLEF--NGPLDGGKAIPIYFGAIIANGNNAI 1230  
 QY 310 FINNISYANSQNL-----GGAIAIDTG-GEISL----- 336  
 Db 1231 LNVNTKLTASHLTGTVAIDINIGAGNLFTIDASVGVTLNANONINERARDSVLYSLN 1290  
 QY 337 -----SAEKGITTFQGNRTSLPLFNGIHLNNAKPLQARNGCSIEF 379  
 Db 1291 TGVGVNNIILAAVLAPAGADEGIVYFNG-----VNLGVNGSVAGTANIDGGGNKF 1344  
 QY 380 -----YDPTSEAD---GSTQNLINGDPKKNKEYGTILFSGEKSILANDPRDKSTIPON 430  
 Db 1345 NTLIIYNAVITTDVNLGEGIONLIN---KNADFTSSTAAGAIQINDATYTIIDANGN 1401  
 QY 431 VNLSAGIIVKEGAETVYSFTQSPGSHLVDLGTKLASKEDIAITGL----- 479  
 Db 1402 LNPAGNIOPAHDAOLV--LONSSGNDRTITIGANIDPNDBGIYLLNSVYAGKLT 1459  
 QY 480 -----AIDISLSSSTAIVIRANTANKOISYDSIEL----- 512  
 Db 1460 AGKTFGAHKLTILFKAGDGCSTAGT-TFNTNIVLIDTGLEGATTAANYVLFNDAY 1518  
 QY 513 -ISPTGN--AYEDLRKNSQTFPPLSLFEPGAGSVYTAGDFLPVSPHYFGQNMKLAFT 569  
 Db 1519 QLTOTGNIGGFLDFPNAKNGAV--TLNNNVVAVAGAVONTG-----TNNGLTIVL 1565  
 QY 570 GTGN-----KYGEFPWDKINRKPRPEKGNLVPNTLNGAVNVRLMOVQETHAS 619  
 Db 1566 GASNLNRVNGIAMLKVG-----AGNV--TAKGGRKYG--EIQGTGTN 1605  
 QY 620 SLOTDGLWIDGIGNEFHVASASDNIYRHNSSGYVLSV-----NE 661  
 Db 1606 TLT-----LPAHFNLGVS-----IKKTGGQALKLFMNGSGSVGVTAANSVGD 1650  
 QY 662 TTPKHHTSNAFSQLFSRDKDYAVSNNEYRYLDLSLYOYTSIGNIFRYASRNPVAVGT 721  
 Db 1651 ITTAGATFSFAS-----VNAKGTATLGG-----TTSFANF-----TNTGA 1686  
 QY 722 LS-----RRELONPLMIFHFLCAYGHATNDKTDYANFPMVNSMNNMCMAIEGSGSM 775  
 Db 1687 VTLAKSITSFAKN-----VTAISFVANSATINFSN-----SLAFNSNI-----TGGGTT 1731  
 QY 776 LVEFNGRFLFGAIPPMKQLVAYAGDEKE-----TTADGRFNSGL----- 819  
 Db 1732 LTLGAN-----QVYTTGTGSFTDITLMTFTFGAAGKSGNLTILKSGSTLDS 1778  
 QY 820 ---TSISVPLGIRFEKLAISQVLY-----DEFS 846  
 Db 1779 GVTSLALVYATNFDDNNNSIPDKTYVISAETAGLKPSTKENVKITINNDNRFVDFTFD 1838  
 QY 847 -----YIPDIFPKDPSCEALVTSGD 867



```

QY 563 NMKLAMTGTGKVGEEFMDKINYPREKEGNL/PNIIIMGNAVVRSLMOYQETHASSIQ 622
Db 761 H-----KLN-----NLLDGGTTFVKELGDTTENGKTKIE 789
QY 623 TDRLGLWIDIGNPFVBASEDNIERYRHNSGGYLVSNNEITPKHYTSM--AFSOLFSDK 680
Db 790 -----GKSLIQISNNYTDHVESADNNGTLEFVNTD 820
QY 681 DYAVSNNEYRYTGLSYLYQTTSIGNIFRYASRNPNVNGIL-----SR 724
Db 821 PIVTTLKOGAYEFGVLKQVILSGPNIV---FNEIGNVGIHGIANSISFENASLGTS 876
QY 725 RFLQDN--PIMFHFLCAYGHATNDKMDYANFNPVKNKSMRNCAIECGSMPLLYENG 782
Db 877 LFLPSGTPLDVLTIKTSVNGT---VDNFPAPIVVSGIDS-----MINNG 919
QY 783 RLFGGAIPEMKLQIVYAYOGDFKETTADGRFRSNGSLTISVPL-----GIRFEK 832
Db 920 QII-----GDKNIIIA-----LSIGSDNSITVANNATLXSGIRTRK 954

RESULT 7
OMP_RICJA STANDARD; PRT; 1656 AA.
AC 006653:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 01-OCT-2000 (Rel. 40, last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxId=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein romp B of Rickettsia
RT japonica."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB003681; BAA20138.1; -
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.
FT DOMAIN 528 533 32 KDA BETA PEPTIDE.
FT POLY-GLY.
SQ SEQUENCE 1656 AA; 168097 MW; 3132a69c9d5999f CRC64;

```

Query Match 3.6%; Score 175; DB 1; Length 1656;  
 Best Local Similarity 20.7%; Pred. No. 0.0091;  
 Matches 192; Conservative 124; Mismatches 318; Indels 292; Gaps 47;

```

QY 11 VFSITLISVPT-----SISA-----TTISLPEDSFHDSQNAERSYV-----QAGDYVS 57
Db 388 IVALTQNSNFETDGFNLAAGVTPVDITLT--GNFTGDANNPGNTAGVTFPAANGTLAS 445
QY 58 LTGDSVSNVNSALNKAFCFNVTSQSVTFAGNH-----HGLYFNINISSG 101
Db 446 ASADANAAYTNNTIAIENS---GVGVQVLSGTHAELRLGAGSVFKLADTVINGKYNQ 502
QY 102 TTKGAVLCCODPOATAFSGFSTLSFQSPDIKEQGLYSKNAMLNNVYVFEQNO 161
Db 503 TVLVGVGLAA---GATLIDGSAFTT-----GDIGNGGGAALQISITLAND----- 544
QY 162 SKTGGAIISGANVTVGVYDVSFYQNAATGGAHSSG-----PLDIANVO----- 208
Db 545 -ATRTLTGLGANI--ISANGTINFOAN---GGTIKLTSTONNTIVDDLAIAIDQGTGV 598
QY 209 --AEIRFPAONTAKKSGGALYSD---GDIDIDONAVYLFRENEALT--TAIGKGAVCCL 261
Db 599 DASSLTNRQTLTIGTIGITIGANNITLQGFNIGSKTTLANGVAINELVIGNNGS----- 654
QY 262 PTSGSSTPVPPIVTFESDNKQVFEERNHSIMGGAIFYARKLISSGPPLFINNISYANSON 321
Db 655 -----VQFAHNTYLTTRTNNAAGCKIIFNPVYN-----NNTTLAAGTN 693
QY 322 LGGA-----IAIDTGEISLSAEKGIITFOGRTSLPLNG----- 357
Db 694 LGSANPLAEINFGSKGARADTVLVNGEVMYATNITTT--DAVNGSVFVAGAKNIYSG 752
QY 358 -IHLONAKFPLKLABNGCTIEF-----YDPTSEADSGTOLINGDPKNEYGTILFS 411
Db 753 TVGGOQGNKFTVALDNGTYKFLGNATFNONTITIAINST--LOISGN-----YRADPLAS 806
QY 412 GEKSLANDPDPFKSTIPQNVNLSAGYLVIRKEAEVYVSKFTQSPSHLYLDLGTKLASK 471
Db 807 AD---GTGIYFVNTGPIVTLNKAQVPAVNAIKQIVS---GPGNVVYNEIGN--AGN 856
QY 472 EDIAITGLAIDIDLSSTSAAY-----IKANRANKO----- 503
Db 857 YHGAMT---DTIAFENSISGAVLFLPSGIPFENDAGNTIPTLTISVGENAEBSVPSV 912
QY 504 -ISYTDSEI---ELISPTGNAYE-DLNRMSQTPPLSLIEPAG-----GSVTVTAGDF 551
Db 913 IVSGVDVYIADGVYIGQNNITVGLGSDNGIYNATTLVAGIGTINNOSTVLISG-- 970
QY 552 LPVSPH--YGFQGNKWLAMTGTGNKVGEEFMDKINYPREKEGNL/PNIIIMGNAVVR 608
Db 971 VPNTPGVYVGL-----CTG--IGASKRKQVTFETDYNNLNIATNTTINDGVTV- 1018
QY 609 SLMOYQETHASSLOTDLGLMIDGI-GNFFH-----VSASEDIRRHNSGGYLVSNNEIT 663
Db 1019 -----TTGGIAGGAGTDFDKITPLGVSNGANVRFAFDGIFNSNTSMIVT 1064
QY 664 PKHYTSMASFQSLFRDQDYAVSNNEYRYTGLSYL-----YOYTTSS-----L 704
Db 1065 TK-----ANNGYTYTLGNAPVAGNIGSDTIVYASVRFKSGNNGAGLK 1105
QY 705 GNIFRYASRNPNVNGILSRRLQNPMLTFHFLCAYGHATN-----DMKTDYANFPYVK 758
Db 1106 GNISOVYIDFTYVLMGIVNSNVI-----LGSTTAINGKIDLLTNLTFLFAGGT 1153
QY 759 NSRRNRCMAIECGSMPLLYFENGRL 784
Db 1154 STWGNNT-SIE-----TTLTLANGNI 1173

RESULT 8
120K_RICRI STANDARD; PRT; 1300 AA.
AC P14914;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, last sequence update)
DT 01-OCT-2000 (Rel. 40, last annotation update)

```

DE 120 KDA SURFACE-EXPOSED PROTEIN.  
GN P120.  
OS *Rickettsia rickettsii*.  
OC Bacteria: Proteobacteria: alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiinae; Rickettsia.  
OX NCBI\_Taxid=783;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-R:  
RX MEDLINE=90136087; PubMed=2515418;  
RA Gilmore R.D. Jr., Joste N., McDonald G.A.:  
RT "Cloning, expression and sequence analysis of the gene encoding the  
RT 120 kD surface-exposed protein of *Rickettsia rickettsii*.";  
RL Mol. Microbiol. 3:1579-1586(1989).  
CC -1 FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS  
CC RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
CC -1 SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A  
CC S-LAYER WITH HEXAGONAL SYMMETRY.  
CC -1 DISEASE: RICKETTSIA RICKETTSII IS THE CAUSATIVE AGENT OF ROCKY  
CC MOUNTAIN SPOTTED FEVER (RMSF).  
CC -1 MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPITOPES  
CC CONFERRING ANTIGENICITY TO THE PROTEIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X16353; CAA34402.1; -  
DR PIR: S07575; S07575.  
DR Antigen: Glycoprotein; S-layer.  
FT CARBOHYD 7 7 N-LINKED (POTENTIAL).  
FT CARBOHYD 66 66 N-LINKED (POTENTIAL).  
FT CARBOHYD 86 86 N-LINKED (POTENTIAL).  
FT CARBOHYD 103 103 N-LINKED (POTENTIAL).  
FT CARBOHYD 147 147 N-LINKED (POTENTIAL).  
FT CARBOHYD 268 268 N-LINKED (POTENTIAL).  
FT CARBOHYD 330 330 N-LINKED (POTENTIAL).  
FT CARBOHYD 375 375 N-LINKED (POTENTIAL).  
FT CARBOHYD 415 415 N-LINKED (POTENTIAL).  
FT CARBOHYD 424 424 N-LINKED (POTENTIAL).  
FT CARBOHYD 430 430 N-LINKED (POTENTIAL).  
FT CARBOHYD 436 436 N-LINKED (POTENTIAL).  
FT CARBOHYD 444 444 N-LINKED (POTENTIAL).  
FT CARBOHYD 515 515 N-LINKED (POTENTIAL).  
FT CARBOHYD 547 547 N-LINKED (POTENTIAL).  
FT CARBOHYD 593 593 N-LINKED (POTENTIAL).  
FT CARBOHYD 655 655 N-LINKED (POTENTIAL).  
FT CARBOHYD 698 698 N-LINKED (POTENTIAL).  
FT CARBOHYD 710 710 N-LINKED (POTENTIAL).  
FT CARBOHYD 799 799 N-LINKED (POTENTIAL).  
FT CARBOHYD 800 800 N-LINKED (POTENTIAL).  
FT CARBOHYD 826 826 N-LINKED (POTENTIAL).  
FT CARBOHYD 844 844 N-LINKED (POTENTIAL).  
FT CARBOHYD 861 861 N-LINKED (POTENTIAL).  
FT CARBOHYD 879 879 N-LINKED (POTENTIAL).  
FT CARBOHYD 920 920 N-LINKED (POTENTIAL).  
FT CARBOHYD 926 926 N-LINKED (POTENTIAL).  
FT CARBOHYD 1116 1116 N-LINKED (POTENTIAL).  
FT CARBOHYD 1128 1128 N-LINKED (POTENTIAL).  
FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).  
FT CARBOHYD 1146 1146 N-LINKED (POTENTIAL).  
FT CARBOHYD 1211 1211 N-LINKED (POTENTIAL).  
SQ SEQUENCE 1300 AA; 132801 MW; E09E52C3647243D CRC64;

Query Match 3.6%; Score 173.5; DB 1; Length 1300;  
Best Local Similarity 20.1%; Pred. No. 0.0079;  
Matches 166; Conservative 102; Mismatches 302; Indels 257; Gaps 41;

QY 25 SATTSLTPEDSEPHGDSOMERSYNOAGDYSL-----TGDVSIAND--NSALNK 74  
Db 415 NGTVKFLGNATFNNGNTTAA--TLQIGNTADCVASADGTGIVFVNTGPIVTLNK 473  
QY 75 ACF-----NYSGSVTFPAGNHGLEYFNISGTTKEAV-----108  
Db 474 QAAPVNAKQITVSGPVGNYVINEIGNAGNHGAVDTIAFENSLGAVFLPGIIPNDA 533  
QY 109 -----LCCDPQATARFSGSTLSFI-----QSPGD-----IKE 137  
Db 534 GNMPFLIKSTVGKTKAKGDVPSVVLGVDSVADQVIGDONNIVGLGSDNGIIVN 593  
QY 138 QGCLYSKALMLNNVYVREQNO-SKTGGAISGANVTIVG-----NDSVSFYON 188  
Db 594 ATTLXA--GISTLNN-----NOGVTVLGGVDPNTPGVYGLGTGIGASKRFVPTFD 644  
QY 189 AATFGAIIHSSGPLQIVNQ-----AETRAQNT--AKNGSGALXSDPDIDIDONAVY 240  
Db 645 YNNGINATINATINDGVTVTGGIAGIGFDKITTGSVNGNGNVRADILNSTSMIG 704  
QY 241 LFRENEALTTAIGKG--GAVCCUPTSGSTPVPPIVPSDKKQVFEERNHSMGGA---I 295  
Db 705 TTKANNGTVTYILGNAPVGN-----GSDTPVASVRTG-----SDSGAGLOGNI 749  
QY 296 YARKLSSISGPTLFINNISYANSQNL--GGAIAIDTGEI-----SLSAEKGTTFQGN 348  
Db 750 YSQVIDRGV-----YNLGIVSNILGGGTALIN--GKIDVNTNLTPTASGSTWGN 800  
QY 349 ---RTSLPFLNG---IHLQNAKFLQARNGCSIEFYPTISEADGSQLNNGPKN 401  
Db 801 TSEITVTLTLANGNIGHIVILEGAQ-----VNTTGTGTTIKVO--DMAN 842  
QY 402 KEYNGTILFSGEKSILANDPDEFKSTIPQ-----NVNLSAGYLVYK-----EGA 444  
Db 843 ANFSGTQY-----TLIOGARFNCTLSPNFAVYGSNRPYNSILIRANODVYITRTNNA 898  
QY 445 EYTV-----SKTQSPGSHVLDLGTKLASKREDIAITGLAIDISLSSSTAAYIKAN 498  
Db 899 ENVTYNDIANSFPGAGG--VDQNTVTFVNTATYANNLLAKNSANSANFEGALVTD 955  
QY 499 TA-----KQISVYDSIELISPTGNAYEDLR-MRNSQFFPLSLSPGAGSVTYAGDFLP 553  
Db 956 TSAITTVQDLDAKDIO--KOLGRIQALYLTGPTAEAMAGPAGA--ISAAVAAGDEAI 1012  
QY 554 VSPHYGFGQWKKLAWTGTGNKVGEEFWDKINY--KPREKEGNLV-----596  
Db 1013 DNYAYG-----IWAKEFYDHAQSKKGLAGYAKATTGVYIGLDTL 1053  
QY 597 --PNIILGNVNVNRSLSMOVETHASSLQTDRLGLMIDG-----IGNFHVASAS-- 641  
Db 1054 ANDMLMIGAAIGITKTDIKHQDYKKGDKTD-----VNGFSFSLYGAQDLVNFPAOGSAIF 1109  
QY 642 -----EDNIRYRHNSGGYLVSNNETPRKHYTSMAFSQLSFSDKOV 682  
Db 1110 SLNOYKKNSQRYEPDANG---NMSKQIAGHYDMTGTGMLTVGYDR 1153

RESULT 9  
ID OMPB\_RICRI STANDARD; PRT: 1654 AA.  
AC Q53047;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).  
GN OMPB.  
OS *Rickettsia rickettsii*.  
OC Bacteria: Proteobacteria: alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiinae; Rickettsia.  
OX NCBI\_Taxid=783;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-R:  
 RX MEDLINE-92167802; PubMed-1724278;  
 RA Gilmore R.D. Jr., Cieplik W. Jr., Pollicastro P.F., Hackstadt T.;  
 RT "The 120 kilodalton outer membrane protein (OMP B) of *Rickettsia*  
 RT *ricketsii* is encoded by an unusually long open reading frame:  
 RT evidence for protein processing from a large precursor.";  
 RL Mol. Microbiol. 5:2361-2370(1991).  
 RN [2]  
 RP SEQUENCE OF 279-1654 FROM N.A.  
 RC STRAIN-R:  
 RX MEDLINE-90136087; PubMed-2515418;  
 RA Gilmore R.D. Jr., Joste N., McDonald G.A.;  
 RT "Cloning, expression and sequence analysis of the gene encoding the  
 RT 120 kD surface-exposed protein of *Rickettsia rickettsii*.";  
 RL Mol. Microbiol. 3:1579-1586(1989).  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X16353; CAA34403.1. -  
 KW AntiGen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1 1333 32 KDA BETA PEPTIDE.  
 FT DOMAIN 1181 1188 POLY-THR.  
 FT SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;  
 SQ  
 Query Match 3.6%; Score 173.5; DB 1; Length 1654;  
 Best Local Similarity 20.1%; Pred. No. 0.011;  
 Matches 166; Conservative 102; Mismatches 302; Indels 257; Gaps 41;  
 QY 25 SATTISTPDSRHGDSQNRBSYNOAGDYSL-----TGDYSISNV-----NSLNRK 74  
 DB 769 NGTIVKFLGNATNGMTTAAANS--TLQIGGNVADCVASADGIGVIEFVNTGPIVTLNR 827  
 QY 75 ACF-----NTSGSVTFAGNHHGLYFNINISGTRKEGAV----- 108  
 DB 828 QAPAVNAKQITVSGPGVYINIEGNAGNHGAVDTIAFENSLGAVYFLPRGIPEND 887  
 QY 109 -----LCCDDPQATRRPSEFSLSTF-----QSFED-----IKE 137  
 DB 888 GNTMPLTKSTVGNKTAAGDPVSVVLGVDVYADGVQVQDNNTVGLGSDNGIIV 947  
 QY 138 QGGLYKKNMLNNTVVRREONO-SKTGKGAISGANVTIVG-----NTYSVSYTON 188  
 DB 948 ATTLVA--GISTLNN-----NQGIVTLLSGVNPMPGVYGLGTGIGASKRFQVFTTD 998  
 QY 189 AATFGAIIHSSGFLQIAVNO-----AEIRPAQNT-----AKNSGALYSDDIDIDONNAVY 240  
 DB 999 YNNLGNINATINATINDGVYTTGTGIGIGEDGKITLGVSAGNKNVRPADSILNNSMIG 1058  
 QY 241 LRENEALTTAIGK--GAVCCLPISGSSSTPVPVIFPSDKOLVFERNHSIMGGA--I 295  
 DB 1059 TTRANNCTVYILNNAFENI-----GSDSDPVASVRFPT-----SDSGAGLGQNT 1103  
 QY 296 YAKRLSSSGPLFLFINNISYANSQNL--GGALAIIDTGGI-----SLSEKTIIFQGN 348  
 DB 1104 YSOVIDFT-----YNGDIVNSNIIILGGGTAIN--EKIDLVTLTLFASGTSWGN 1154  
 QY 349 ---RTSLPFG--IHLQNAKFLKLARNGCSIEFYDPIITSEADSGTQNLINIDPKN 401

DB 1155 TSIEETTLTLANGNIGHVILEGAQ-----VNTTGTGTTIKVQ--DNAN 1196  
 QY 402 KEYGTTLFSGEKSLLANDPDEFKSTIPQ-----NNLSAGYLVK-----EGA 444  
 DB 1197 ANFSGTQTY-----TLIOGARFNGTCLGSPNFAYTGSNRFVNSLIRANODYVTRTNA 1252  
 QY 445 EYVY-----SKTFQSGSHLVLDLGTKLASKEDIAITGLAIDIDLSSTSTFAVKAN 498  
 DB 1253 ENVTYNDIANSPFGAG--VDQNTVTFVANNTAAYNNLLAKNSANSANFVGAIVTD 1309  
 QY 499 TA-----NKQISVTDISIELISPTGNAYEDLR--MRNSQTFPLSLTEPGAGSVTYAGDFLP 553  
 DB 1310 TSAIITVQDLAKDIO--RQIGNRGALRYLIGTPEFAEWAEPAGA--ISAAYVAGDEAI 1366  
 QY 554 VSPHYFGQWMLKAMWTGKNGVGEFWDKINY--KPEKRGNTLV----- 596  
 DB 1367 DNYAAG-----IWAKPFYIDAHOSKKGGLAGYKAKTGVVIGLDTL 1407  
 QY 597 --PNIILMGNVNRSLMQVQETHASSLQTRGLMIDG-----IGNFHVAS-- 641  
 DB 1408 ANDNLMIAGAIIGTKYDKIKHODYKKGDKTD--VNSFSLSLGAQQLVKNFPAQSGAIF 1463  
 QY 642 -----EDNIRYRHSNGGYVLSVNNETPPKHYTSMASFSDKDY 682  
 DB 1464 SLNQYKRKKSQRFFPDANG---NMSKOIAAGHYDNMTGGLTYGYDY 1507  
 RESULT 10  
 ID OMPB\_RICTY STANDARD; PRT; 1645 AA.  
 AC P66989;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
 DE (CONTRAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).  
 GN SLIP.  
 OS *Rickettsia typhi*.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsia.  
 OX NCBI\_TaxID=785;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WILMINGTON;  
 RX MEDLINE-94040787; PubMed-8224886;  
 RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;  
 RT "Cloning and sequence analysis of the gene encoding the crystalline  
 RT surface layer protein of *Rickettsia typhi*.";  
 RL Gene 133:129-133(1993).  
 RN [2]  
 RP PARTIAL SEQUENCE.  
 RC STRAIN-WILMINGTON;  
 RX MEDLINE-92114896; PubMed-1370573;  
 RA Ching W.M., Carl M., Dasch G.A.;  
 RT "Mapping of monoclonal antibody binding sites on CDBR fragments of  
 RT the S-layer protein antigens of *Rickettsia typhi* and *Rickettsia*  
 RT *proszekii*.";  
 RL Mol. Immunol. 29:95-105(1992).  
 RN [3]  
 RP IDENTIFICATION OF CLEAVAGE SITE.  
 RX MEDLINE-92104668; PubMed-1729180;  
 RA Hackstadt T., Messer R., Cieplik W., Peacock M.G.;  
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer  
 RT membrane protein of *Rickettsia*: identification of an avirulent  
 RT mutant deficient in processing.";  
 RL Infect. Immun. 60:159-165(1992).  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A

CC S-LAYER WITH HEXAGONAL SYMMETRY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC

DR EMBL: L04661; AAB48987.1; -

KW AntiGen; S-layer; Transmembrane; Cell wall.

FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.

FT TRANSMEM 1354 1645 32 KDA BETA PEPTIDE.

FT CONFLICT 1415 1429 MEMBRANE ANCHOR (POTENTIAL).

FT CONFLICT 657 657 H -> N (IN REF. 2).

FT CONFLICT 842 842 V -> I (IN REF. 2).

FT CONFLICT 1071 1071 G -> A (IN REF. 2).

FT CONFLICT 1306 1306 G -> S (IN REF. 2).

FT SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

Query Match 3.6%; Score 172; DB 1; Length 1645;

Best Local Similarity 20.1%; Pred.No.0.014; Indels 342; Gaps 53;

Matches 185; Conservative 124; Mismatches 271;

QY 21 DMSLSATTT--ISLPEDSFH-----GDS--ONAESYVWQAGVSLTGVSISVNDN 69

DB 687 DTFALAGCTNGSAPLSNHFATKANGSILHIGKGVLYANNI--TTTD--ANVGS 741

QY 70 SALNKACFNVTSGSVTFAGNHGLYFNN--ISSGTRKEGAVLCCDPQATARFSGFSTLS 127

DB 742 LHRSGSGTSIVSGTV---GGQGLKLTNLLINDGT-----TVKFLGDTTFN 784

QY 128 FLOSPDDEKQGLYKRNALMLNNTVVRPQOSKT-----KGGK--- 168

DB 785 ----GGTKIEG--KSLQISSNITTDHIESADNTGLEFVNTDPTVTLNKQGAFFG 835

QY 169 ----ISG--ANYTI--VGN-----YDSVIFYONAAFGAIIHSSG--PLQIAVNOAET 211

DB 836 VLKQWVSGSGNIAFMEIGNGVAHATAVDISF--ENASLGASLFLSGTFLDVLTKSTV 894

QY 212 RFAQNTAKNGSGALYSDGDIDIDONAYVLFRENEALTTAIGKGAVCCLPTSGSST-PV 270

DB 895 --GNGVDFNENAPILVYSGIDSMINGVIGDQKNIIALSLGSDNFI---TVNSMTLYA 948

QY 271 PIYTFEDNKQVVERHNSIMG--GATYAKLSISSGCP-----TLFNNNSYAN 318

DB 949 GRTTKTNGTV---TLSGGIPNNPGTLYG--LGLENGDPRLKQVTFPTDYNNGSLII 1001

QY 319 SONL-----GGAIDDTGGEISL-----SAEKGT 342

DB 1002 ANVTATINDVTLTGTGAGTDPGKTKITLSSINGNANVKFVDRFSPHTSIATVETKANQGT 1061

QY 343 TFFQGN-----RTSLPFLNGIHLLONAKFLKQAR--NGCSIEF--YDPTTSEAD--- 388

DB 1062 VYTLGNALVGNISGSDIP--VASVRFNGDSGVGLQGNINHQNIDFTYNTLIINSDVILG 1120

QY 389 -GSTQAINNDGPKNKREYTGITLFSGEKSLANDPRDKRSTIPQAV--NLSAGVLYIKGAEV 446

DB 1121 GGTTAINGEID---LLTNMLIFANGSTWGNNTSLSTLT--NVSNGNWQGVIVAGAAQV 1174

QY 447 TV-----SKFTQSPGS-----HLVLD 462

DB 1175 NATTTGTTTIKIDNANANPSSGTQTTLLILOGGARFNGTLCARPFYDTGANNIFPKYELIRD 1234

QY 463 -----LGTKLIAS-----KEDIAITGLAIDLSLSSSTAAVYKAN 498

DB 1235 ANQDYVLTFRNDVLYAVTAVGNSAIANAPGVHQNTAI-----CLSTEDTAAY----- 1282

QY 499 TANKQISVDSIE-----LISPTGNAYEDLPMRNSQ--TFPLLSLEPGA----- 540

DB 1283 --NNMLIARDSSDVAVFIGAIAITDTGAAVAATVNLNDTOKTODLLGNRLGLRYLSNSETA 1340

QY 541 --GGSVT--VTAGDFLPVSPHYGFGNMMKLAMTGTGNKYGEFFMDKINRK--PAPREKGNL 595

DB 1341 DVGSGETGAVSSGDEALIDVSYG-----VMAPFPNIARQDKKGL 1381

QY 596 V-----PNILGNANVNRSLMOVQETHASSLQTD--RLGLWIDIG--- 633

DB 1382 AGYKATAGVVGDLDTLANDLNMLGAIAGITKTIDIKHODKRDKTDIKGLSFLGACQ 1441

QY 634 --NRFVNAS-----EDNIRYRHNNGCYLVSNNETTPKHYTSMAP--SQLRSRQK 680

DB 1442 LVKNFPAQGSATFTLKNVSKSQRYFFDANG--KNNKQIAGNVDNITFGGMLMGFYDI 1498

QY 681 D-----YAVSNNE 688

DB 1499 NALQGVLYTFPMAGLSYLLKSSNE 1520

RESULT 11

ID YDBA\_ECOLI STANDARD; PRT; 2003 AA.

AC P3366; P76087; P76088; P76856; P76857; P76859;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE HYPOTHETICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.

GN YDBA OR B1401/B1405.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

ON NCBI\_TaxID=562;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-K12;

RX MEDLINE=97251357; PubMed=9097039;

RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,

RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,

RA Makino K., Miki T., Mizoduchi K., Mori H., Mori T., Motomura K.,

RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,

RA Sempel G., Seki Y., Sivasubaram S., Tagami H., Takeda J.,

RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;

RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome

corresponding to the 28.0-40.1 min region on the linkage map.";

RL DNA Res. 3:363-377(1996).

RN [3]

RP SEQUENCE OF 464-2003 FROM N.A.

RC STRAIN-K12;

RX MEDLINE=92190338; PubMed=1665988;

RA Moszer I., Glaser P., Danchin A.;

RT "Multiple IS insertion sequences near the replication terminus in

Escherichia coli K-12.";

RL Biochimie 73:1361-1374(1991).

CC -1- SIMILARITY: TO S. TYPHIMURUM ORF NEAR CYSG (AC P25928).

CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION: THE GENE CODING FOR

THIS PROTEIN IS INTERRUPTED BY A HYBRID IS20/IS30C ELEMENT

BETWEEN AMINO ACIDS 839 AND 840.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: AE000237; AAC74483.1; ALT_SEQ.
DR EMBL: AE000237; AAC74487.1; ALT_SEQ.
DR EMBL: D90778; BAA1509.1; ALT_SEQ.
DR EMBL: D90778; BAA15080.1; ALT_SEQ.
DR EMBL: D90779; BAA1881.1; ALT_SEQ.
DR EMBL: X62680; NOT_ANNOTATED_CDS.
DR Ecogene; EG11307; ydda.
KW Hypothetical protein.
FT CONFLICT 489 I -> V (IN REF. 2).
FT CONFLICT 495 I -> V (IN REF. 2).
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;

Query Match 3.58; Score 170; DB 1; Length 2003;
Best Local Similarity 18.6%; Pred. No. 0.024;
Matches 203; Conservative 139; Mismatches 337; Indels 414; Gaps 53;

QY 20 FDTSLSAT-----ISLPEDSFHDSQNAERSYV-----Q 51
DB 1053 YDSSLANFTLVGAAYEHGAINDQGVININVTGQAFYNDGTGTIVNGTICTFQVQC 1112

QY 52 AGDYVSLT-----GVSISNVDSALNK--ACFNYSVTSVTFAGNHGTYFNIS 99
DB 1113 SGNEYNTDDEFTSLYTGDTITRSGETVILNKSAVTDKLAGNVVNSGLISQDI-TV 1171

QY 100 SG---TRKEGAV--LCCQDPQATARFSGFSTLFSIQPGDIKEGCLYKRNAL-----M 148
DB 1172 SGLLENTSGIINLVKLDKGAIVKKNAGVMTNNVDVSGILLNNAGMTAQITMAGADSS 1231

QY 149 LLNN--YVREFEQ-----QSKTGGAISGANV--TIYGNADV 183
DB 1232 LVNNTGTINKIVQAGVFNNSGVTSGRMSAGVFNQTDGALIRGALGTAVANNEGT 1291

QY 184 -----SFYONAAVF-----GGAISGPIQAVNOAE 210
DB 1292 WNLGSSSEGNNTGMLFVNNSAFNNRGEFILDNDKNAVHINOSTVINTGHMIS----- 1346

QY 211 IREFAONTAKN--GSGALYSDGDI-----DONAVYRENEALTTAIGKG 255
DB 1347 -NSHNGAVNMWGGNGRFINDGTIDVASAKLVASANNAGQNAFEMQNDGVINFHDSA 1405

QY 256 GAVCCLEPTSGSSTFVPTVTSNDKOLVERNHSIM--GGAIYARKLSSSGPTLFIN 312
DB 1406 SAV-----KVTSHN--FLQNDGIMNISGAV-----AMEGDKNAQLVN 1443

QY 313 NISYANSONLGAIAIDT-----GEISISAER-----GITTFQ 346
DB 1444 N-----GTINLCTAGTTDTTGMIQMOLDANATADAVIENNGTINIRANDSEAFSVLGTIGHV 1499

QY 347 GNRTSLPLNGI--HLLONAKFLKLOARNG--CSIEFYD-----PTSE 386
DB 1500 VNNGTVIADGVGSGLLIKQGDSTINVEGMNGNNSSEVHYGDTLPDVPKPTVSTSG 1559

QY 387 AD--GSTQLININDPKRKETGTLFSEKSLANDPDEKSTIPQNNVLSAGIYVIR--- 441
DB 1560 SDEAGSGMNNLNG-----YVGTNVNVSAGKGLKVNNS 1592

QY 442 -EGAEVVSFETGSPSHVLVDLGTKLIAKEDIAITGLAI-DIDSLSSSTAIVTAMT 499
DB 1593 MNGVEIWTG-FT-----AGTADTVYSPDNVVGSMLTBDADITSTSVWTAAGST 1641

QY 500 -ANKQISVTDSTIELISFTGNAYEDLR-----MNSQTFPLP----- 534
DB 1642 DASGNVDVTS-----KNAVTDVATDASVNDIAKALDAGYTNNELFTSLNVTGAELN 1694

QY 535 -SLEPGAGSVYTAGDFLVSPIHSGQWKLAMTGTGKVGFEFFDKINYK--PRPEK 591
DB 1695 SALKQVSGSQATTVFARVLSNFSMLAD-----AAPKVGGLAFNVYVAGDPRREL 1747

QY 592 EGNLVPIILGNVNVNVSLSQVQETHASSLQTRDGL-WIDGIGNEFHVASSEDNIRVRHN 650

```

```

DB 1748 GNNREYDML-----ALKRTIDLSQSQTMSL--EYGIARLDGDG--AQKAGDNCV----- 1792
QY 651 SGGYVLSVNNETITPKHTSMASFQSFQSKDKDYAVSNNEYRYLQSYLYQYTSIGNIFY 710
DB 1793 TGGY-----SOFGLKHMSPDNG-----MNNNALRY 1820
QY 711 ASRRPNVNGILSRPLQNPMLIFHPLCAVGH-----ATNDKTDYANPVMYKNSWRNWCW 766
DB 1821 DVHNLDSRSI-----AFGNTKKTADTDVKKQYLEFRS----- 1853
QY 767 AIECGSMPLLVFENGRLFOGA-----IPFKLQLVYAYOGDFEKTAD--GRRESNGL 819
DB 1854 -----EGANTFESEGLKVTYPYAGVKLRHTLEGGYQGBRANAGDENLMMNSGSE 1900
QY 820 TSISVPLGIREKALASQDVLYIDFSFYIPIFKRDPSCBAALVYISDSWLVPAANSRH 879
DB 1901 TAVDSIYGLK-----DYGKDGMSASATLEGPN--LSYAKSORT 1939
QY 880 AFVSGGTGRYHFN 892
DB 1940 ASL-AGAGSOHFN 1951

RESULT 12
YpJA_ECOLI
AC P52143; P76610; P77017; P77019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 98.4 KDA PROTEIN IN ALPA-GABD INTERGENIC REGION (F949).
GN YpJA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGI655;
RX MEDLINE=97426617; Pubmed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPRAIN=K12;
RA Alba H., Bada T., Fujita K., Hayashi K., Honjo A., Horluchi T.,
RA Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K.,
RA Kasai H., Kashimoto K., Kim S., Kimura S., Kitagawa M.,
RA Kitakawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H.,
RA Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Tagami H., Takemoto K., Wada C.,
RA Yamamoto Y., Yano H.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).
CC - SIMILARITY: STRONG, TO BORDETTELLA PERTACTIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: U36840; AAA79815.1; ALT_SEQ.
DR EMBL: AE000350; AAC75695.1; -.
DR EMBL: D90889; BAA16514.1; ALT_INIT.
DR EMBL: D90890; BAA16518.1; ALT_INIT.
DR Ecogene; EG13213; ypJA.

```





```

QY 385 -----SEADSTQTLNINGD-----PKNKEYTG-----TILFSGEKS 415
DB 1017 AIKGGGTTNNALETAKKS-QFTSNGDISINNGENAHBGAQFQKQKTYINAGDL 1075
QY 416 LANDPRPKFTIPONVLSAGLYL-----IKGAEVYVSKFTQSP 455
DB 1076 TLAQATDTHSESONVNGSANLKGFTPESKDYGGCFNAGTTHSKQQTAKVGTINGSQ 1135
QY 456 GSHVYLDLGRKL-----IASKEDIAITGL-AIDIDLSSTSAVAIKANTANKQI----- 504
DB 1136 G--TELNAGNLLTQGTHTLSSEDDIALNATNKVDLOSASSEHTE--KGNVLSGGVQAGF 1190
QY 505 --SVTDSIELISPTGNAYEDLRMRNSOTFPLLSLPGA-----GGSV----- 544
DB 1191 GKMTDDASSVNLGSGQFALIGKODEKS---VSREGTINNSGMLTNGNSVHLQGAQVN 1247
QY 545 -----TVTAADF-LPVSPHYFGQNMKIAMTGTGNKVGEFFMDKINYPKPR--PEKEGN 594
DB 1248 SKDTQLTSQSGDIEITSAOSTDYKNNM-----GFDIG-FNGCKTNTPEKYEEXKPAT 1299
QY 595 LVPRILMGNAVNRSLMOVQ-----ETHASSLOTDRGLMIDG-----IGNFFH 637
DB 1300 SIHNIGKLLVNEVDQKTSQONATLETGTLTINSKRDLLSGANVTADSVTGVNGSLN 1359
QY 638 VSASED-----NIRYRH-----NSGGVYL-----SVNNEI 662
DB 1360 IASQKESDRHVTVGVANGVYHTNDPKSSOVNKTAKAGGSILEKTIKDTIDSGIKSTDAI 1419
QY 663 TPKHYTSMASQLEFSRDQDAVSNNEYRMYLGSYLYQYTTSLGNIFRYASRNPNVNGIL 722
DB 1420 SDR-----YNSLSSTIADTKTGISDETK---AKIDGFGKVGNGIK-----NIVAG-- 1461
QY 723 SRNFLONPLMIFHLFCAYGATN--DMKTDYANFPYKNS---WRNNOMALECGGSML- 776
DB 1462 -----AEGHTANADIKVTHVDNDAVTKTSLTSLNNDLSLVNGSTKTLFG 1505
QY 777 --LVFENGRLFOGALPEMKQLVAYVQGDKEETAD 810
DB 1506 AEIVSQGGVYDLGG-SSVKLENT---EGHHEAGAD 1537

RESULT 14
HLTA_SERMA STANDARD; PRT; 1608 AA.
AC P15370;
DB 01-APR-1990 (Rel. 14, Created)
DB 01-APR-1990 (Rel. 14, Last sequence update)
DB 01-NOV-1990 (Rel. 16, Last annotation update)
DE HEMOLYSIN PRECURSOR.
SHLA.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RC STRAIN=SN8;
RX MEDLINE=88257037; Pubmed=3290200;
RA Poole K., Schiebel E., Braun V.;
RT "Molecular characterization of the hemolysin determinant of Serratia
marcescens.";
RL J. Bacteriol. 170:3177-3188(1988).
CC -I- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
DEFINED.
CC -I- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
REOIRRES SHLB FUNCTION.
CC -I- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -I- SIMILARITY: TO P. MIRABILIS HEMOLYSIN (HPMA).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; M22618; AAA50323.1; -.
DR PIR; A28182; A28182.
KW Hemolysis; Toxin; Outer membrane; Signal.
FT SIGNAL
FT CHAIN 31 1608 HEMOLYSIN.
SQ SEQUENCE 1608 AA; 165078 MW; D669B476FE7DAD51 CRC64;

Query Match 3.4%; Score 164.5; DB 1; Length 1608;
Best Local Similarity 22.2%; Pred. No. 0.038;
Matches 148; Conservative 104; Mismatches 242; Indels 173; Gaps 35;

QY 42 QNERSYVNOAGDYSLT---GDVSI---SNVDSN-----ALNKAQFNVTSQVTF 87
DB 853 ENAAGVDHLAADTASTTTKTIDYGVNIGANVDYSATRYPERAVGAALADATGVINDI 912
QY 88 GNHGLEYENNISSGTYKEGAVLCCQDPQATARESGFSTLSF-IQSPGDIKQGLY--SK 144
DB 913 G---GIGAPNVGIDIGAQGSSEKSSSQAVVSSVQAGSIDINAKGEVRDQTOYQASK 969
QY 145 NALMLL-----NNVVFEDQNSK-TKGGA-----ISGANVT-----VGNDSVSYQ 187
DB 970 GAVNLTADSHRSEAAANROEDQSDHTGSGAGVRYTTTGSDLTVDADKEGQTORSSNS 1029
QY 188 NAAFTGGAHSSGPIQAVNOAEIRFAONTAKNGSG--ALYSDGIDIDONAVLREN 245
DB 1030 QAYT--GSDAANGINNVKKDAI--YGTALNGRKRKTAVNAGDRLQDAS---DKS 1082
QY 246 EALT-----TA-----IGKGAVCLPTSGSSTPYPIYTFSDNKOVLPERNH 287
DB 1083 ESRSGFNVKASAKGFTADSKNFAGGFGGTHNGESSSTAQVGNISGQGVLEKAGRDL 1142
QY 288 SIMGGCAIVAKLSSSG----- 305
DB 1143 TLGGTDVKSQGDVSLSGANKVALQAAESTQTRKESKLSGNIDLGAGSSDSKEKTGNLSA 1202
QY 306 GPLTFINNISYANSONGALAIPTGEISLSAE-KG--TTFQGNRTS-----LPLNG 357
DB 1203 GGAFDIAKVNESATERGATITASD--GKVLISANGKDDALHQAAGVSGSALAEKNG 1260
QY 358 IHLQNAKFLKLOARNCSIEFYDPTTSEADGSTQLING---DPRNKEYTGTILFSGE 413
DB 1261 GILLESK--NEQHKRWML-----GIRANAKGGOTFKRQDAGKVDPTGKQDHT-IGAGL 1313
QY 414 KSLANDPRDKFTIPONVNLISAGLYLKEGAETVSKFTQSPGSHLVLDL-----GTLK 467
DB 1314 K-VGEVQD--KTHANTGTITAGDVTLSNG-----KDTRLAGARVADSVQKVGDLH 1364
QY 468 IASKEDIAITGLAIDIDS--LS-----SSSTAAYIKANTANKQISVDSIE-----LISP 515
DB 1365 VESKRQVE-NGVAYVDVAGLSHNDPSSITSKLSKVTGTRYAKVKEKLEAGNKKADA 1423
QY 516 TGNAYEDLRMRNSOTFPLLSLPGAQSVTVTAGDLPVSPHYFGQNMKIAMTGTGNKY 575
DB 1424 TTDKYNVSARR-----LDPQDPTGAVSFSKA-----EKGVLPLPAPPEKPR 1465
QY 576 GEFVWDK 582
DB 1466 QGPLMDR 1472

RESULT 15
YI09_YEAST STANDARD; PRT; 995 AA.
AC P40442;
DB 01-FEB-1995 (Rel. 31, Created)
DB 01-FEB-1995 (Rel. 31, Last sequence update)

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 2, 2001, 03:31:02 ; Search time 119.05 Seconds  
(without alignments)  
1024.655 Million cell updates/sec

Title: US-09-446-677B-6

Perfect score: 4814

Sequence: 1 MRFSLCGPPLVFSLLLSVF.....IECRHARNYNNICSGKFRF 922

## Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 100%

Listing first 45 summaries

## Database :

SPREMBL\_16:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.podent:\*  
12: sp.unclassified:\*  
13: sp.vertebrate:\*  
14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4814	100.0	922	2	0924H9	0924H9 chlamydia p
2	4794	99.6	922	2	0929C5	0929C5 chlamydia p
3	4790	99.5	922	2	09K1Y9	09K1Y9 chlamydia p
4	1516	31.5	926	2	P71135	P71135 chlamydia p
5	1441	29.9	928	2	092398	092398 chlamydia p
6	1434	29.8	930	2	09R866	09R866 chlamydia p
7	1432	29.7	930	2	092393	092393 chlamydia p
8	1400.5	29.1	936	2	092898	092898 chlamydia p
9	1400.5	29.1	936	2	09J542	09J542 chlamydia p
10	1393	28.9	928	2	09R865	09R865 chlamydia p
11	1353	28.1	914	2	086163	086163 chlamydia p
12	1344	27.9	928	2	086164	086164 chlamydia p
13	1344	27.9	949	2	09K299	09K299 chlamydia p
14	1258	26.1	847	2	P71132	P71132 chlamydia p
15	1249	25.9	772	2	09R871	09R871 chlamydia p
16	1242	25.8	839	2	P77792	P77792 chlamydia p
17	1202	25.0	846	2	P71133	P71133 chlamydia p
18	1137.5	23.6	973	2	092896	092896 chlamydia p
19	1137.5	23.6	995	2	09K2A1	09K2A1 chlamydia p

20	1063.5	22.1	1407	2	092899	092899 chlamydia p
21	1060	22.0	1276	2	09JRM2	09JRM2 chlamydia p
22	1007	20.9	841	2	0923A1	0923A1 chlamydia p
23	951.5	19.8	1013	2	084879	084879 chlamydia t
24	901	18.7	987	2	09PL45	09PL45 chlamydia m
25	796.5	16.5	712	2	09R873	09R873 chlamydia p
26	774.5	16.1	867	2	09PL41	09PL41 chlamydia p
27	727	15.1	878	2	084882	084882 chlamydia t
28	720	15.0	445	2	09R867	09R867 chlamydia p
29	678	14.1	649	2	P71134	P71134 chlamydia p
30	629.5	13.1	359	2	09J5K6	09J5K6 chlamydia p
31	610	12.7	1723	2	092812	092812 chlamydia p
32	610	12.7	1723	2	09R859	09R859 chlamydia p
33	610	12.7	1732	2	09K2C1	09K2C1 chlamydia p
34	581	12.1	946	2	092880	092880 chlamydia p
35	581	12.1	946	2	09R860	09R860 chlamydia p
36	577	12.0	1609	2	09R858	09R858 chlamydia p
37	577	12.0	1609	2	092605	092605 chlamydia p
38	576.5	12.0	427	2	09R870	09R870 chlamydia p
39	574	11.9	938	2	092883	092883 chlamydia p
40	569.5	11.8	934	2	092882	092882 chlamydia p
41	569.5	11.8	934	2	09J5E7	09J5E7 chlamydia p
42	569.5	11.8	952	2	09K2A5	09K2A5 chlamydia p
43	561	11.7	1016	2	084880	084880 chlamydia t
44	559	11.6	983	2	09PL44	09PL44 chlamydia m
45	558.5	11.6	1672	2	09PJY2	09PJY2 chlamydia m

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	922 AA.
0924H9	0924H9			
AC	0924H9			
DT	01-MAY-1999 (TrEMBLrel. 10, Created)			
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)			
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)			
DE	OUTER MEMBRANE PROTEIN 6 PRECURSOR.			
GN	OMP6.			
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI_TaxID=83558;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-VRIJ10;			
RA	Hierne K., Boesen T., Daugaard L., Knudsen K., Madsen A., Christiansen G., Birkelund S.;			
RT	"Chlamydia proteins containing the GAT-repeat belong to a subfamily of autotransporting pathogenicity factors."			
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AJ13035; CAB37075.1; -			
DR	InterPro: IPR003357; -			
DR	InterPro: IPR003358; -			
DR	Pfam: PF02385; OMP.1.			
DR	Pfam: PF02415; DUF145; 1.			
KW	Signal.			
FT	SIGNAL. 1			
SO	SEQUENCE 922 AA; 100362 MW; 0DD025E118ACB1CA CRC64;			

Query Match 100.0%; Score 4814; DB 2; Length 922;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 922; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRFSLCGPPLVFSLLLSVFDTLSATTISLPEDSFHGSQNAERSYNNQAGVYSLTG 60
DB	1	MRFSLCGPPLVFSLLLSVFDTLSATTISLPEDSFHGSQNAERSYNNQAGVYSLTG 60
QY	61	DVTSINVDNSALNKACPNVTSGSVTFAGNHGILYFNNTSGTTKEGAVLCCDDQATARF 120
DB	61	DVTSINVDNSALNKACPNVTSGSVTFAGNHGILYFNNTSGTTKEGAVLCCDDQATARF 120

```

QY 121 SGFTLSFISPGDIKQGLYSKNALMLNNYVFEQNSKTKGAGISANTYIVGN 180
DB 121 SGFTLSFISPGDIKQGLYSKNALMLNNYVFEQNSKTKGAGISANTYIVGN 180
QY 181 DSVFYONAAATFGGAIHSSGPIQIAVNOAETRFQONTAKNGSGGALYSDGIDIDQNAVY 240
DB 181 DSVFYONAAATFGGAIHSSGPIQIAVNOAETRFQONTAKNGSGGALYSDGIDIDQNAVY 240
QY 241 LFRNEALTTAIGKGAIVCCLPITSGSSTPVPYTFSDNKOLVFERHNSIMGGAIVARKL 300
DB 241 LFRNEALTTAIGKGAIVCCLPITSGSSTPVPYTFSDNKOLVFERHNSIMGGAIVARKL 300
QY 301 SISSGPTLFINNIVYANSONLGAIADYGGELISAEKGTITFOGNRTSLPLNGIHL 360
DB 301 SISSGPTLFINNIVYANSONLGAIADYGGELISAEKGTITFOGNRTSLPLNGIHL 360
QY 361 LQNAKFLKQARNGCSIEFDPTISEADGSTOLINNDPKNKETGTITLSEKSLANDP 420
DB 361 LQNAKFLKQARNGCSIEFDPTISEADGSTOLINNDPKNKETGTITLSEKSLANDP 420
QY 421 RDEKSTIPQNVNLSAGLYLKEGAETVSKFTQSPGSHLVLDLTKLIASEKEDIATIGLA 480
DB 421 RDEKSTIPQNVNLSAGLYLKEGAETVSKFTQSPGSHLVLDLTKLIASEKEDIATIGLA 480
QY 481 IDIDSLSSSTAAYIKANTANKOISVTDIELISPTGNAYEDLRMRNSQTFPLSLBGA 540
DB 481 IDIDSLSSSTAAYIKANTANKOISVTDIELISPTGNAYEDLRMRNSQTFPLSLBGA 540
QY 541 GGSVTVAGDFLPVSPHYGQGNKMLAMTGTGNKVGFEFMDKINRKPRPEKGLVFNIL 600
DB 541 GGSVTVAGDFLPVSPHYGQGNKMLAMTGTGNKVGFEFMDKINRKPRPEKGLVFNIL 600
QY 601 WGNANVRSLMQVQETHASSLQTDRLGIMIDIGNFFHVSASEDNIRYRHNSGGYVLSVN 660
DB 601 WGNANVRSLMQVQETHASSLQTDRLGIMIDIGNFFHVSASEDNIRYRHNSGGYVLSVN 660
QY 661 EITPKHTYSMAFSOLFSDKDYAVSNNEHYMGLSYLYOTTSIGNIFRYASRPNVNVG 720
DB 661 EITPKHTYSMAFSOLFSDKDYAVSNNEHYMGLSYLYOTTSIGNIFRYASRPNVNVG 720
QY 721 ILSRFLONPIMTFHFLCAVGHATNDKTDYANFPMVKNRNNCMALIECGSMPLLVFE 780
DB 721 ILSRFLONPIMTFHFLCAVGHATNDKTDYANFPMVKNRNNCMALIECGSMPLLVFE 780
QY 781 NGRLEFOGAIPEFMKQLVYAGDEKETTADGRPSNGSLTISVPLGIREKIALSDVL 840
DB 781 NGRLEFOGAIPEFMKQLVYAGDEKETTADGRPSNGSLTISVPLGIREKIALSDVL 840
QY 841 YDSEFSTIPDLFRKDPCEALVYISGDSWLVPAHVSRAHAFVSGTGRYHENDYTELLCR 900
DB 841 YDSEFSTIPDLFRKDPCEALVYISGDSWLVPAHVSRAHAFVSGTGRYHENDYTELLCR 900
QY 901 GSTICRPHARNYINNGSKFRF 922
DB 901 GSTICRPHARNYINNGSKFRF 922

```

```

RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger U., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE-20330349; PubMed-10871362;
RA Shihai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL; AE001585; AAD18163.1; -
DR EMBL; AP002545; BAA98215.1; -
DR InterPro; IPR003357; -
DR InterPro; IPR003368; -
DR Pfam; PF02385; OMP; 1.
DR Pfam; PF02415; DUF145; 1.
SQ SEQUENCE 922 AA; 100457 MW; DFE2AB6333AB031C CRC64;

Query Match 99.68; Score 4794; DB 2; Length 922;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 919; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRFSICGPEPLVFSLTLLSVFDTSLATTISLTPEDSFPHGDSQNAERSYNOAGPVSYLTG 60
DB 1 MRFSICGPEPLVFSLTLLSVFDTSLATTISLTPEDSFPHGDSQNAERSYNOAGPVSYLTG 60
QY 61 DVSISVNDNSALNACFENVTSGSVTFAGNHGGLYFNNISSGTTKEGAVLCCQDPQATARF 120
DB 61 DVSISVNDNSALNACFENVTSGSVTFAGNHGGLYFNNISSGTTKEGAVLCCQDPQATARF 120
QY 121 SGFTLSFISPGDIKQGLYSKNALMLNNYVFEQNSKTKGAGISANTYIVGN 180
DB 121 SGFTLSFISPGDIKQGLYSKNALMLNNYVFEQNSKTKGAGISANTYIVGN 180
QY 181 DSVFYONAAATFGGAIHSSGPIQIAVNOAETRFQONTAKNGSGGALYSDGIDIDQNAVY 240
DB 181 DSVFYONAAATFGGAIHSSGPIQIAVNOAETRFQONTAKNGSGGALYSDGIDIDQNAVY 240
QY 241 LFRNEALTTAIGKGAIVCCLPITSGSSTPVPYTFSDNKOLVFERHNSIMGGAIVARKL 300
DB 241 LFRNEALTTAIGKGAIVCCLPITSGSSTPVPYTFSDNKOLVFERHNSIMGGAIVARKL 300
QY 301 SISSGPTLFINNIVYANSONLGAIADYGGELISAEKGTITFOGNRTSLPLNGIHL 360
DB 301 SISSGPTLFINNIVYANSONLGAIADYGGELISAEKGTITFOGNRTSLPLNGIHL 360
QY 361 LQNAKFLKQARNGCSIEFDPTISEADGSTOLINNDPKNKETGTITLSEKSLANDP 420
DB 361 LQNAKFLKQARNGCSIEFDPTISEADGSTOLINNDPKNKETGTITLSEKSLANDP 420
QY 421 RDEKSTIPQNVNLSAGLYLKEGAETVSKFTQSPGSHLVLDLTKLIASEKEDIATIGLA 480
DB 421 RDEKSTIPQNVNLSAGLYLKEGAETVSKFTQSPGSHLVLDLTKLIASEKEDIATIGLA 480
QY 481 IDIDSLSSSTAAYIKANTANKOISVTDIELISPTGNAYEDLRMRNSQTFPLSLBGA 540
DB 481 IDIDSLSSSTAAYIKANTANKOISVTDIELISPTGNAYEDLRMRNSQTFPLSLBGA 540
QY 541 GGSVTVAGDFLPVSPHYGQGNKMLAMTGTGNKVGFEFMDKINRKPRPEKGLVFNIL 600
DB 541 GGSVTVAGDFLPVSPHYGQGNKMLAMTGTGNKVGFEFMDKINRKPRPEKGLVFNIL 600
QY 601 WGNANVRSLMQVQETHASSLQTDRLGIMIDIGNFFHVSASEDNIRYRHNSGGYVLSVN 660
DB 601 WGNANVRSLMQVQETHASSLQTDRLGIMIDIGNFFHVSASEDNIRYRHNSGGYVLSVN 660
QY 661 EITPKHTYSMAFSOLFSDKDYAVSNNEHYMGLSYLYOTTSIGNIFRYASRPNVNVG 720
DB 661 EITPKHTYSMAFSOLFSDKDYAVSNNEHYMGLSYLYOTTSIGNIFRYASRPNVNVG 720

```

```

QY 721 ILSRFLQNPMLTFHFCAGHATNDKTDYANPMPYKNSMRNNCAIECGGSMPLLYFE 780
DB 721 ILSRFLQNPMLTFHFCAGHATNDKTDYANPMPYKNSMRNNCAIECGGSMPLLYFE 780
QY 781 NGRIFOGAIFPMKQLQVYAYAGDFKETTADGRFNSGLTSISVPLGIREKTLASQDVL 840
DB 781 NGRIFOGAIFPMKQLQVYAYAGDFKETTADGRFNSGLTSISVPLGIREKTLASQDVL 840
QY 841 YDFSFSYIPDIFRKDPSCFALVIGSDSWLVPAHVSRAHAFVSGTGRYHFNHYTELLCR 900
DB 841 YDFSFSYIPDIFRKDPSCFALVIGSDSWLVPAHVSRAHAFVSGTGRYHFNHYTELLCR 900
QY 901 GSTECRPHARNYINNGSKRFF 922
DB 901 GSTECRPHARNYINNGSKRFF 922

RESULT 3
Q9K1Y9 PRELIMINARY; PRT; 922 AA.
AC 09K1Y9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
GN CP0770.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN SEQUENCE FROM N.A.
RP STRAIN-AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman G., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AE002237; AAF38570.1; -.
DR TIGR: CP0770; -.
DR InterPro: IPR003357; -.
DR InterPro: IPR003368; -.
DR Pfam: PF02415; DUF145; 1.
SQ SEQUENCE 922 AA; 100480 MW; 3442106888AB0FA6 CRC64;

```

```

Query Match 99.5%; Score 4790; DB 2; Length 922;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 918; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MRSLSGEPVLELTLSVVDTSLSATTSITLPEDSHGDSQAHERSYNQADVYSLTG 60
DB 1 MRSLSGEPVLELTLSVVDTSLSATTSITLPEDSHGDSQAHERSYNQADVYSLTG 60
QY 61 DVSISVNDNALKKACFNVTSGSVTFAGNHGGLYFNINISGTTKEGAVLCCODPOTARF 120
DB 61 DVSISVNDNALKKACFNVTSGSVTFAGNHGGLYFNINISGTTKEGAVLCCODPOTARF 120
QY 121 SGFSYSLFISQPEDIRKQGLYSKNAFMLNLYVREONQSKTKGALISGANVTIVGN 180
DB 121 SGFSYSLFISQPEDIRKQGLYSKNAFMLNLYVREONQSKTKGALISGANVTIVGN 180
QY 181 DVSFVONATPGGATIHSSGPIQIAVQAETIRRAQNTAKKSGGALYSDDIDIDONAVY 240
DB 181 DVSFVONATPGGATIHSSGPIQIAVQAETIRRAQNTAKKSGGALYSDDIDIDONAVY 240
QY 241 LFRNEALTTAIGKGAIVCLPTSGSSTPVPIVTFSDNKLQVFRNHSIMGGAIVARKL 300
DB 241 LFRNEALTTAIGKGAIVCLPTSGSSTPVPIVTFSDNKLQVFRNHSIMGGAIVARKL 300

```

```

DB 241 LFRNEALTTAIGKGAIVCLPTSGSSTPVPIVTFSDNKLQVFRNHSIMGGAIVARKL 300
QY 301 SISSGPTLFINNISTYANSONLGALAIIDTGGETSLSAEKGTTTPOGNNRSLPLNGIHL 360
DB 301 SISSGPTLFINNISTYANSONLGALAIIDTGGETSLSAEKGTTTPOGNNRSLPLNGIHL 360
QY 361 LONAKFLKQARNGSTIEFPDITSEADGSTOLININDPNKREYTGILISGEGSLANDP 420
DB 361 LONAKFLKQARNGSTIEFPDITSEADGSTOLININDPNKREYTGILISGEGSLANDP 420
QY 421 RDEKSTIPQVNVLSAGLYVIRKEGAETVSKETQSPGSHLVLDGTLKJASKEDAIATGLA 480
DB 421 RDEKSTIPQVNVLSAGLYVIRKEGAETVSKETQSPGSHLVLDGTLKJASKEDAIATGLA 480
QY 481 IDIDSLSSTAAVIRKANTANKOISTVDSIELISPTGNAYEDLMRNSQTFPPLSLPGA 540
DB 481 IDIDSLSSTAAVIRKANTANKOISTVDSIELISPTGNAYEDLMRNSQTFPPLSLPGA 540
QY 541 GGSVTVAGDFLPSPHYGQGNKFLAWGTGKNGVGEFMDKINYPKPEKEGMLVPIIL 600
DB 541 GGSVTVAGDFLPSPHYGQGNKFLAWGTGKNGVGEFMDKINYPKPEKEGMLVPIIL 600
QY 601 WGNANVRSIMQVQETASSLQTDRLGIMIDIGNFFHVSASEDNIRYRHNSGCVLSVN 660
DB 601 WGNANVRSIMQVQETASSLQTDRLGIMIDIGNFFHVSASEDNIRYRHNSGCVLSVN 660
QY 661 ETPKHTYSMAFSQLEFRDQDAVSNNEYRMVLSYLYQYTTSLGNIFFRYASRNPVNVG 720
DB 661 ETPKHTYSMAFSQLEFRDQDAVSNNEYRMVLSYLYQYTTSLGNIFFRYASRNPVNVG 720
QY 721 ILSRFLQNPMLTFHFCAGHATNDKTDYANPMPYKNSMRNNCAIECGGSMPLLYFE 780
DB 721 ILSRFLQNPMLTFHFCAGHATNDKTDYANPMPYKNSMRNNCAIECGGSMPLLYFE 780
QY 781 NGRIFOGAIFPMKQLQVYAYAGDFKETTADGRFNSGLTSISVPLGIREKTLASQDVL 840
DB 781 NGRIFOGAIFPMKQLQVYAYAGDFKETTADGRFNSGLTSISVPLGIREKTLASQDVL 840
QY 841 YDFSFSYIPDIFRKDPSCFALVIGSDSWLVPAHVSRAHAFVSGTGRYHFNHYTELLCR 900
DB 841 YDFSFSYIPDIFRKDPSCFALVIGSDSWLVPAHVSRAHAFVSGTGRYHFNHYTELLCR 900
QY 901 GSTECRPHARNYINNGSKRFF 922
DB 901 GSTECRPHARNYINNGSKRFF 922

```

```

RESULT 4
P71135 PRELIMINARY; PRT; 926 AA.
AC P71135;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PUTATIVE 98 KDA OUTER MEMBRANE PROTEIN.
OC Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN SEQUENCE FROM N.A.
RP STRAIN-OVINE ABORTION S26/3;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U72499; AAB18188.1; -.
DR InterPro: IPR003357; -.
DR InterPro: IPR003368; -.
DR Pfam: PF02385; OMP; 1.
DR Pfam: PF02415; DUF145; 1.
SQ SEQUENCE 926 AA; 98439 MW; 3E755E2F594750F CRC64;

Query Match 31.5%; Score 1516; DB 2; Length 926;
Best Local Similarity 37.8%; Pred. No. 9.3e-91;

```

Matches 359; Conservative 160; Mismatches 379; Indels 52; Gaps 22;

```

OY 1 MRSLSGFLPLVETLITVSDTSLSATTSITLPEDSPFHGDSQNAERSYNOAGDVYSLTG 60
Db 1 MRSLSKILISLTLPPISEHFSQLAHEVALTDESLIDANGARSPPOSTSTAGGTIVVES 60
OY 61 DVSISNV-DNSALNKACFNTVSGSVTFAGNHGLYFNNTSSGTTKEGAVLCODPOATAR 119
Db 61 DISIVDVGQAAALASSAFVGTADNLTFKGNNSLSTINANAGAPGINVMTADKILTF-- 118
OY 120 FSGFSLTSLFQSGDIKE--QGCLYSKNAMLANNVYREDOQSKTKGAGISANTV 177
Db 119 LDFSESLTSEKCESSLVNTGKGAMKSGGALANLNNASITLFDNYSAENGAGISCKAFSLT 178
OY 178 GNDVSFEYONA-ATFGAISHSGPLQIAVNOAEIRFAONTANGSGGALYSGDDIDQ 236
Db 179 GSKSEISFTTNSAKKAGGALATGIAHLSNOCITIRFSGTAAVN-SGGANVYSRASMIIAG 237
OY 237 NAVLFRRENALTTAIGKGAVALCPLTSGSSTPVPVTFSDNKOLYFERNHSIMGGAIT 296
Db 238 NNHVAFSNNAVSGSSDGGCGAHCSTK-GSA---PFLTRDNKVLTFEEMTSAGKGAIT 293
OY 297 ARKLSSSGGPTLFINN-ISYANSQNGALADTGEISLSAKGTITQGRITSLP-- 353
Db 294 TDKLITLSSGPTAFINKVTHATPK--GGATGIAANGCESLTAEHGDITFDNMLMATQDN 351
OY 354 ---FLNGIHLLQAKFLKLOARNGCSIEFYDPTSEADGTOQINODPKNKRYGTILF 410
Db 352 ATTKRNAINEGKGVNLRASGKITSFYDPTTVEGNADLTLTKAKEDKTYNGRIIF 411
OY 411 SGCK---SLANDPRDKRSTIPQVNVNSAGYLVYKEGAETVSKFTSPGSHLVLDJGTL 467
Db 412 SGCKLEQGAADVADNKTFTPTQITLAAAGELVRSGEVEAKTVQTAGSLIIMDAGTKL 471
OY 468 IASKEDIAITGLIDISLSSSTAIVKANTKQISVDSITELSPITNAVEDLRMRN 527
Db 472 SATTEBATLTNLININPTLDGKR-FAVYDVAAGKNTVTSAGALGVDPDGTKEYNKKL-- 528
OY 528 SQEPPLSLSEPGAGSVYTAGDFLP-----VSPHYGFGNNKMLAMTGTGN-----KVG 576
Db 529 NDPLALGGIQLSGKSVTTGN--VPSHYVGAETHGYGGMNSVSVKDNNSDPKQTQTA 585
OY 577 EFTWDKINKPREPEKGNLVPNLKGNANVRSLSMOYQETHAS-LOTDRGLMIDIGNF 635
Db 586 IFWNKTGYVNERBRAPVLNLSMGSFIDLRSDIOLERSVSIETRRGLWVGSGTGNF 645
OY 636 FHYASASDNRHYRNGSGYVLSVNNELTPKHYTSMASFOLFSDKQOAVSNNEYRWLGS 695
Db 646 FHRDRNENKFRHISGIVLAGATTNTSRDLSVAFCOLFAKDKDVLVSKNANAYAGS 705
OY 696 YLQYTTSLGNIFRYASRNPNVWVGLSRFLONPLMIFHFLCA---YGHATDMKTDYA 752
Db 706 VYQYHVSFKPDRLRN-GRNTCCSGSKSI--PI----FLDAQIITYCHTANMTTSYF 757
OY 753 NFPVYKNSMRNCAIECGSMPLVFEENGRLFOGALPEPKLQVAYOGDEFKETTADR 812
Db 758 DYDEVGSMGNDTLGLTSLVPIPVFSSS-IFDSYAPFAKLOVYVAHODDFEKEPTTEGR 816
OY 813 RFGNSLTSIVPLIRFELASODVLYDFSPSYIPDIRKPOSCSEALVIGSDSWLVP 872
Db 817 VRESPLLNVSPVIGIKFEKLSIGERSAIDLMTYIPDYRHNPSCHMTGLAIDVSWLTT 876
OY 873 AARVSHAEVSGCTGRYHNDYTELRCGSIECRPHARANNYNGSKFRP 922
Db 877 ATPLAARAFIVRAGNHIALTSGVEMFSQGFELRSSRNNVNDLGAKVAF 926

```

RESULT 5  
 Q92398 PRELIMINARY; PRT; 928 AA.  
 AC Q92398;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (ITEMblrel. 10, last sequence update)

DT 01-MAR-2001 (TREMblrel. 16, last annotation update)  
 DE OUTER MEMBRANE PROTEIN PRECURSOR (POLYMORPHIC MEMBRANE PROTEIN G  
 DE FAMILY).  
 GN OMP10 OR PMP\_9 OR CP0306  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae)  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiales.  
 ON NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CW1029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VRI310;  
 RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,  
 RA Christiansen G., Birke Lund S.;  
 RT "Chlamydia proteins containing the GGAT-repeat belong to a subfamily  
 RT of autocatalyzing pathogenicity factors.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBD databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Uuchi K.,  
 RA Shiba T., Ishi K., Hattori M., Kuhnare S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CW1029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman G., Dodson R., Gwin M., Nelson M., DeBoy R., Kolonay J.,  
 RA McManis C., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AE001628; RAD18591.1; -;  
 DR EMBL: AJ133034; CAB37069.1; -;  
 DR EMBL: AP002546; BAA98655.1; -;  
 DR EMBL: AE002192; AAF38163.1; -;  
 DR TIGR: CP0306; -;  
 DR InterPro: IPR003357; -;  
 DR InterPro: IPR003368; -;  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.  
 KW Signal.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 928 OUTER MEMBRANE PROTEIN.  
 SQ SEQUENCE 928 AA; 98332 MW; 58910A8F04F12219 CRC64;

Query Match 29.9%; Score 1441; DB 2; Length 928;  
 Best Local Similarity 36.2%; Pred. No. 7, 5e-86;  
 Matches 348; Conservative 169; Mismatches 373; Indels 70; Gaps 25;

```

OY 1 MRSLSGFLPLVETLITVSDTSLSATTSITLPEDSPFHGDSQNAERSYNOAGDVYSLTG 57
Db 1 MRSLSKILISLTLPPISEHFSQLAHEVALTDESLIDANGARSPPOSTSTAGGTIVVES 60
OY 58 LTGDSISNVND-SALNKACFNTVSGSVTFAGNHGLYFNNTSSGTTKEGAVLCODPOA 116
Db 61 LTGDSVITNAGSPTALTASCFKETTGNLSFGCHGYOFLQNDIDAGAN-----CTFTNT 113
OY 117 TAR----FSGFSLTSLFQSGDIKEQGCLYSKNAMLANNVYREDOQSKTKGAGISAA 172
Db 117 TAR----FSGFSLTSLFQSGDIKEQGCLYSKNAMLANNVYREDOQSKTKGAGISAA 172

```



```
Db 114 AANKLSFGSGFSLIQTNNATGTAIKSTGACSIQSNVSCVFGQNFSDNGALQGS 173
Qy 173 NNTLVGVNDVSFQONAT-FCGAIHSSGPIQIAVNOAEIRFAONTAKNGSGALYDGD 231
Db 174 SISLSLN-PLNTEFAKNKATQKGALYSTGTGITTNNLNSASFSENTAAN-NGAIYTEAS 231
Qy 232 IDIDONAVYFREAREATTIAGKGAACCLPTSGSSFPVPIVFSMDKQVFEENHIMG 291
Db 232 SFTSSNAISF-INNSVTASARGAAYC---SSTSAKPVLLSDNGELNIGNVATIS 287
Qy 292 GCAIYARKLSISSGPPPLFINNISYANSQNLGAIAIDTGEISLSAEKGTTFQGN--- 348
Db 288 GCAIYTNVLVSSGPPPLFNNSAIDTAPLGAIADSGSLSLAGDITTEGTVV 347
Qy 349 -----RTSLPLNGIHLQNAKFLKQANNGSIEFPDITSE--ADGSTQNLIN 396
Db 348 KGASSQTTNRNSINICN-----TNAKIVQLRASQGNITVFYDPTITSITAAISDALNLN 402
Qy 397 GD-PKKKEYGTITLPSGK---SLANDPRDKSTIPONNLSAGYLVIREGAENVYSKF 451
Db 403 GPDLAGNPATOGTIVSGEKLSEAEADNLSTIQDPLTAGGQLSLSGVTLVAKSF 462
Qy 452 TQSPGSHLVLDGTLKASKEDIAITGLAIDISLSSSTAAYIKANTANKQISVTDSIE 511
Db 463 SQSPGSTLMDAGTTL-ETADGITINLVNVSLSKETK-KATLIKATQASQVTVLSSGSL 520
Qy 512 LISPTGATYEDLRNRSQTPPLSLSPGAGSVTVT--ACDFLEVSP-HYFGQNKWLAW 568
Db 521 LVDPGSGVYEDVSNMNPQVSCLTLTADDPANHITDLADPLEKNPIHNGYOGNMAISW 580
Qy 569 ---TGTKNGKGEFFMDINKPREKEGNLVPNILMGNAVNSLMQVGFHASSLOTDTR 625
Db 581 QEDTATSKAATLTWTGTGYNPNPERGTLVANTLMSFVDVNSIQDLVATKYRQSOETR 640
Qy 626 GLWIDIGNEFFHVASADNIRYRNSGCVLVSNNETPKRYHSMASQLESRDKDAVS 685
Db 641 GIMCEGISNEFHSDSTINKGFRHISAGYVGAFTTLASNLNLTAAACQLEKGRDHFIN 700
Qy 686 NNEYRMLVGSYLYQ--TSLGNIFRYASRNPVNVGILSRRLQNPMLTFHFLCAYGHA 743
Db 701 KNRAAFAAASLHQLHATLSSPSLRLTPGSES-----EQP-VLFDAQISYIS 748
Qy 744 TNDKRTYANFPVKNKSRMNCNAIECGSMPLLVFENGRLFGAIFPMLQLVYAAYQD 803
Db 749 KNTMKYITTOAPGESSWMDGCLASSLPHTALSHEGLFHAFFPIVEASYSIHQDS 808
Qy 804 FKE-TTADGRFFSNGSLTISVPLGIFEREKLAISODLVYFSESYIPDIRKDPSCAAL 862
Db 809 FKERNTLVASFDSDGLINVSPIGIFERFSRNERASTYATVYVADYRKNPDCITLAL 868
Qy 863 VIGDSMLVPAHAVSRHAFVSGTGRYHENDYELLCRGSIIECPHARANYININGSKFRF 922
Db 869 LINTSWKTGTNLISROAGIGRAGIFYAFSPNLEVTISNLSMEIGSSRSYNADIGKFOR 928

RESULT 6
ID 09RB66 PRELIMINARY; PRT; 930 AA.
AC 09RB66;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE
DE PROTEIN G FAMILY).
GN PMP_8 OR CP0307.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shitai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
```

```
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CwL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39.
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,
RA Barry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M., and Chlamydia
RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AP002546; BAA98654.1; -.
DR EMBL: AF002193; BAA98164.1; -.
DR TIGR: CP0307; -.
DR InterPro: IPR003357; -.
DR InterPro: IPR003368; -.
DR Pfam: PF02385; OMP; 1.
DR Pfam: PF02415; DUF145; 1.
SQ SEQUENCE 930 AA; 97639 MW; 46A8896761391C09 CRC64;

Query Match 29.88; Score 1434; DB 2; Length 930;
Best Local Similarity 36.64; Pred. No. 2,2e-85;
Matches 344; Conservative 157; Mismatches 395; Indels 44; Gaps 21;

Qy 10 LVPSLTILSVFDDTSL--ATTISLTPEDSFH--GDSQNAEBSYVVOAGDYVSLTGDVSI 65
Db 8 LLISSTLVTPILLISAYYAGADASLPTDSEFGAGSGFTTPESTADANGTNVLVSGNVYIN 67
Qy 66 NV-DNSALNKAACEFNVTSYVFAGNHGLYFNINISGTTKEGAVLACODDPATRFSGFS 124
Db 68 DACKGATLTGCCFETTGDLFTGCKGYSFNTYVDASNGAASTADAKLT--FPGFS 125
Qy 125 TLSFIQSPGDKEQG--CLYSKNAFMLNNVAREFO--NQSRTKGAGISGANVTVGN 179
Db 126 NLSPIAPAGTTVASGKSTLSAGALNTDNGTILFSQNVSEANNGNGALYAKTLAISGN 185
Qy 180 YDSVSPQONA-TFGCAIHSSGPIQIAVNOAEIRFAONTAKNGSGALYSPGDIDIDONA 238
Db 186 TSTTFPNSNAKKLGAIYSSAAISGNTGQVFNANKGTG-GGALGFEASSITQNS 244
Qy 239 YVLFRENEATTTAIGKGAACCLPTSGSSFPVPIVFSMDKQVFEENHNSIMGATYAR 298
Db 245 SLFSGMTA-TDAAGKGAAYCEKGTG---ETPLTISGKNSLFFAENSVYOGGATCAH 299
Qy 299 KLSISSGPPPLFINNISYANSQNLGAIAIDTGEISLSAEKGTTFQGN--RTSLP-- 353
Db 300 GLDLSAAGPPLFSNNRCNTAAGKGAIAIADSGSLSLASNOGDITFLGNTLSTSAPTS 359
Qy 354 FLNGIHLQNAKFLKQANNGSIEFPDITSEADGSTQ-LINIGDPKMK--:LITLFL 410
Db 360 TRNAVLYGSSAKITNLAAAGOSTIFYDPIASNTGTASDVLTLLQDPDSNPLDTSGLTVF 419
Qy 411 SGEKSLANDPR--DFKSTIPONVNLASAGYLVIREGAENVYSKTFQSPGSHLVLDGTLK 467
Db 420 SGEKLSADEKADNFIILKQPLASGLTALKGNVELDVNGTQREGSLMQPQTKL 479
Qy 468 IASKEDIAITGLAIDISLSSSTAAYIKANTANKQISVDSIEFLISPTGNAYEDLRNRN 527
Db 480 KADTEAISTRLKVLVDLSALGNKSVSIEFAG-ANKTITLTSPLVFQDSNGFYESHRTNQ 538
Qy 528 SQFPPLISLSPG-AGSVTVTAQDGLPV---SPHYFGQNMKLAFTGTGN-KVGEFFMDK 582
Db 539 AFTQPLVFTTAATPAASDIYDALITSPVQTPPEPHYGTQGHWEATWADTSTAKSGTMTWT 598
Qy 583 INTKPREKEGNLVPNILMGNAVNSLMQVGFHASSLOTDRLGIDIGNEFFHVASSE 642
Db 599 TGVNPNRERASVVPDLSMASFDIRLQQLMISQANSTIQQRGLMASGTANFPHKXSG 658
```

Oy	643	DNIERRHNSGGVLSVNNEITPEKHTSNAFSOLFSDKDKRYAAVSNNEYAGSLYQYRT	702
Db	659	TNOAERHRHSYGTIGVCGSADESENFTEFACOLFGKDCKLTFIVENTSHNYLASLYIQHA	718
Oy	703	SILGNTERFASRNPNNVNGLISRRFLONPLMIHFCAAGHATNDMKTDYANPPMYKNMR	762
Db	719	FLUGLT-----PMSFSGITMDLANDIPL-LLNAQLSYSTYKNDMDRTRYTSYPEAQGSWT	770
Oy	763	NNCMAIECCGSHPLLVFENGRLFOQAIPPMKLQLYAYAYOGDKEKTADRRRSNGSLTSI	822
Db	771	NNSGALIELGGSSIALTLPRKAPPFQGYFFPLKKQAYYSROQNKESGAEARADDDDLVCNC	830
Oy	823	SVPLIFREKTLASODVLYDFSFYSYIPIEDFRDPCECALTYISGSGMWLPAAVSRSHAFT	882
Db	831	SIPVRIELKEIDEEDKNKFELAIITGVYRKNPSPRSJSLMWSGASWISLCNNLARQAFPL	890
Oy	883	GSGTGGRHYENDYTELLCRGSIETCPRHARNYNINCGRKFEF	922
Db	891	ASAGSHLTLSPHVELSGEAAYELRGSAHIYVNDGCLARSF	930
RESULT	7		
ID	O9Z393	PRELIMINARY;	PRT; 930 AA.
AC	O9Z393:		
DT	01-MAY-1999 (TREMBLrel. 10. Created)		
DT	01-MAY-1999 (TREMBLrel. 10. Last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16. Last annotation update)		
DE	OUTER MEMBRANE PROTEIN 11 PRECURSOR.		
GN	OMP11 OR PMP_8.		
OS	Chlamydia pneumoniae (Chlamydophila pneumoniae).		
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.		
OX	NCBI_TaxID=83558;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CWL029;		
RX	MEDLINE=99206606; PubMed=10192388;		
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,		
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.; Hyman R.W.,		
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";		
RL	Nat. Genet. 21:385-389(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-VRI310;		
RA	Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,		
RT	Christiansen G., Birkelund S.		
RT	"Chlamydia proteins containing the GGAI-repeat belong to a subfamily		
RT	of autoantigenic pathogenicity factors."		
RL	Submitted (JAN-1999) to the EMBL/genbank/DBJ databases.		
DR	EMBL: AE001627; AADI8590.1; -		
DR	EMBL: AJ133034; CAB87068.1; -		
DR	InterPro: IPR003357; -		
DR	InterPro: IPR003368; -		
DR	Pfam: PF02385; OMP; 1.		
DR	Pfam: PF02415; DUF145; 1.		
KW	Signal.		
FT	CHAIN 1 27 POTENTIAL.		
FT	CHAIN 28 930 OUTER MEMBRANE PROTEIN 11.		
SO	SEQUENCE 930 AA; 97669 MW; 46A9B5E3BB913C4C CRC64;		
Query Match	29.7%; Score 1432; DB 2; Length 930;		
Best Local Similarity	36.6%; Pred. No.2.ge-85;		
Matches 344;	Conservative 157; Mismatches 395; Indels 44; Gaps 21;		
Oy	10 LVFSLTLLSVEDFSL--ATTISLRPEDSFH--GDSONAERSYNYOAGDYSLTGVDYSIS	65	
Db	8 LLISTLVTPILLSTATYAGDASTLPTDSFDGAGSTFPKRSTADANGINVYLSCGWYIN	67	
Oy	66 NV-DNSALNKACEFNVTSSGVTPFAGNHGILYFINNISSGTTKEBAYLVCODPOATARFSGS	124	
Db	68 DAGKTAALTGCCFTTETTDLDLFTTGVSYSFTVPVAGSNAAGAATTDALKALT--FTGFS	125	

QY	125	TLTSLFGSGDILKEQG--CLYSKNNMLMLNNVAREPEO---NOSTKSGAIGCAVNTLVN	179
Dd	126	NLSFLAAAGTTVAASCKSTLSLSAGALNTLDNKTILFSPQNVSEANNGAITTKTSLISGN	185
QY	180	YDVSVEFYONAA-TEFGAITHSSGPLOIAVNOAIEIRFAONTAKNGSGGALYSDDGIDIDOMA	238
Dd	186	TSSTIFTSNSAKKLGALYSSAASISGNTQGLVFEMNNKGETG--GGALGFEAASSITONS	244
QY	233	YVLFENEALTTAIGKGAACCLPTSGSTVPTVYTSNDKQLVFERNNHSMGGALYAR	298
Dd	245	SLTFEGGNA-TEAARKGAAYICENTKG---ETPLLTITSGNKSILTEFANSSVTQGCALCAH	299
QY	299	KLSISGCPPTLEFINNIYSAANSQNLGCAIADTGEISLSAEKGTITTEQGN---RTSLP--	353
Dd	300	GIDLSAAGPTLFSSNNRCGNTFAAGGALAIAMDSSLSLSANQDGIPTLGLNTLSTSAPTS	355
QY	354	PLNGIHLQNMKFLKLOARNGSCSIEFYDPTISEADGSTO-LININGDKRNK--EYTGILF	410
Dd	360	TNNAIYLGSSAKITNLRAAQOSIYFEDPIASNTTGGASDVLTINQPPNSPLDYSCTIVE	419
QY	411	SGEKLADPR---DFKSTIQONWLSAGIYIKEGAETVYSKTSQPSGLVLDLGTKL	467
Dd	420	SGEKLSAEAKAAMDFTILQOPALASGTLALGNVELDNGETQEGSTLLMQPETKL	479
QY	468	IASKEDIAITGLAIDIDLSSSSTAAYIKANTAKKOISVTDSIELISPTGNAVEDLBRN	527
Dd	480	KADTEAISITLKLYVDLSLBEKGSYSITTAG-AKKTTITLSPVAFODSSGNYFSHTINQ	538
QY	528	SOTFPLLSLEBEG-AGGSYTVTAGDFLPV---SPHYGFGNMKLAMTGTGN-KVGEFFMDK	582
Dd	539	AFTQGLVETATATAASDIYIDALLTSPVQTEPEHYGQGHWEATWADTSPAKSGTMTVWT	598
QY	583	INAKRPERKEGNLVPNITMGAAVNRSLMOQEETHASSLOTRDLRLNIDGIGNPFHVSASE	642
Dd	599	TGYNNPERRASVYDPDLSMASFTDIRTLQIMTSQANSIYQORLMSAGTANFEHKRSG	658
QY	643	DNIRFHNHSGGVLSVNNNEIIPKTHISMAFSQLSRBDYAVSNNKRYMLGSLYQYTT	702
Dd	659	TNOAFRHSSTGYIVGSGADESENFSAFQOLGFKDKDLFIYENTSHNTLASLYLOHRA	718
QY	703	SIGNTFRYASRNPVNVNGILSRFLQNLMTIHFELCAVGAHTNDKMTDYANFPMVKNSMR	762
Dd	719	FLGGI-----PMPSEFSITDMKLDP-L-ILNAQLSYSTKNDMDRYTSYPAQGSWT	770
QY	763	NNCMALIEGGSPFLIYFENGRLPFGALJPFMKLOLVYAVQGFETTDGRRFSGSLTSI	822
Dd	771	NNSGALEIGGSLALPKPEADPFQGYRPFLLFQAVYSQQNFKSGSGEARFDDGDLYNC	830
QY	823	SVPLAIFREKTLASDVLVYDSEFSYPTDIFIKRDCSEALVYISGDSWLVPAAHYSRAHY	882
QY	883	GGGCTGRHFNDYTELLCRGSIECPRHANVYINGSKFRF	922
Dd	831	SIPVGIRLEKISEDEKKNFEISLAYIGDVYKRNRSRPSILWVSCASWTSICKNLARQAF	890
Dd	891	ASAGSHLTLSPHVELSGEAAVELKGSAAHYNVWDGLGRYSF	930
RESULT	8		
Q9Z898			
AC	Q9Z898;	PREDIMINARY;	PRT; 936 AA.
DT	01-MAY-1999 (Tremblrel, 10, Created)		
DT	01-MAY-1999 (Tremblrel, 10, Last sequence update)		
DT	01-MAR-2001 (Tremblrel, 16, Last annotation update)		
DE	POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY, PMP-7.		
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).		
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
OX	NCBI_TaxID=83558;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN-CWL029;		

RX MEDLINE-99206606; PubMed-10192388;  
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";  
 RL Nat. Genet. 21:385-389(1999).  
 DR EMBL; AE001627; AAD18589.1;  
 DR InterPro; IPR003357;  
 DR InterPro; IPR003368;  
 DR Pfam; PF02385; OMP; 1.  
 DR Pfam; PF02415; DUF145; 1.  
 SQ SEQUENCE 936 AA; 100079 MW; 8881D78A53D194EC CRC64;

Query Match 29.1%; Score 1400.5; DB 2; Length 936;  
 Best Local Similarity 35.9%; Pred. No. 3.4e-83;  
 Matches 338; Conservative 166; Mismatches 391; Indels 47; Gaps 21;

QY 9 PLVFSLLTSLVPTSLSATYISLTPEDSFHGDSQNAERSYVOAGDYVSLTGPVSISSND 68  
 DB 14 PLVFSLLTSLVPTSLSATYISLTPEDSFHGDSQNAERSYVOAGDYVSLTGPVSISSND 69  
 QY 69 --NSALKACFNTSGSVTFAGNHGLEYENNISGTTKEGAVLCCODPQATPARSGFSTL 126  
 DB 70 ALGIPLASGCFLEAGDGLTFQGNQHALKFAFINAGSS-AGTVASTSAADKNLTFNDFSRL 128  
 QY 127 SFIQSGDIKE---QGCIVSKNALMLNNVYFEONOSTKRGALISGAVTVGNDSV 183  
 DB 129 SIISSPSLLSPGQCALSKVGNLSLTGNSQIIFTQNFSSDNGCINTKFFLLSGTSQFA 188  
 QY 184 SFYQNA---TFGGAIHSSGPLOIAVNOAEIFRQONTAKSGSGALYSDDIDIDONAVY 240  
 DB 189 SFRNGAFNGKGGVYFATCTITENSFGVFSQNLAK-GSGGALXTSTNCSTITDFQV 247  
 QY 241 LFEENALTAICKGAACCLPTSGSTPVPVYTFSDNKOLVERNHSINGGALYARKL 300  
 DB 248 IFGNSMMEWAQQAQGAICCTTDTK-----VTLGKMKLSFTNMTALYGGALISGLKY 301  
 QY 301 SIISSPSLLSPGQCALSKVGNLSLTGNSQIIFTQNFSSDNGCINTKFFLLSGTSQFA 356  
 DB 302 SIISSPSLLSPGQCALSKVGNLSLTGNSQIIFTQNFSSDNGCINTKFFLLSGTSQFA 361  
 QY 357 GHLLQNAKFLQARNGCSIEFYDPTS--EADGSTOLINDGPKNK--EYTGTLIFSG 412  
 DB 362 ALNITDIAKTYSIRATGSGIYFYDPTNGTAASTDTLNLADANSEIEYGAIVFSG 421  
 QY 413 EKSLAND---PRPKSTIQONVLSAGLYIKCAEYVYSKFTQSPGSHVLVDLGRKLA 469  
 DB 422 EKLSPEKALIAAVTSTIRPAVLARGDLVLRDGVYTFEDLQSPGSRILMDGGTLLSA 481  
 QY 470 SKEDIAITGLAIDISLSSSTAAYIKAMTANKOISVTOSIELISPTGAYEDLRMRNSQ 529  
 DB 482 KEANLSINGLAVALSLDGTNRKA-LKTEADANISLSTGITALIDTBSGYEENHNLKAS 540  
 QY 530 TPELLSL-EPGAGGSVTVYAGDLPV---SPHYFGQNMKLAN-TGTGNKVGGEFMDKIN 584  
 DB 541 TYPLLELTAGAGTITLGLSLTLTLOEPETHGYGQNMQLSMANNTSSIGSINMTRTG 600  
 QY 585 YKRPKEKGLVNNILMGANVNRSLMOVETIASSLQTDRCGLMDIGNEFHYVASEDN 644  
 DB 601 YIPSPKRSKSLPLNSLMGNFIDIRSLINOLETSSGSEPERELMLSGIANFERYDSMPTFR 660  
 QY 645 IRRHNSGGVYLVNNELPRKHTYMAFSQLEFRDQYAVANNEMMYLGSYLYOYTTSL 704  
 DB 661 HGRHISGATLIGITTTAEDQDLTFACQLARDRKNTLTKGKHGDTTGASLYFHTHGL 720  
 QY 705 GNIFRY---ASRPNVNVGILSRRLONPLMIFELCAVGHATNDKTDYANFPVKNS 760  
 DB 721 FDIANFLMGATRAP---WVLSISQIIPLS-FDAKFSYLHNDMKTYTTONSIIKGS 775  
 QY 761 WRNNCAIECGSMPLLYENGRLEFGALIPMKLQLVAYAGDFKETTADGRFSNGSLT 820  
 DB 776 WRNDACADGASLP-FVISVPLKKEVEPFVQYIYAHQODFYERHAGRAFNKSELI 834

QY 821 SISVPLIGREFKALISODVLYDFSEFYIPDIFRRKPSCEALVYSGDSWLVPAHVSRIA 880  
 DB 835 NVEIPDGVTEROSKSKGYDTLMTILLDAYRNPCCQSLASDMMAYGTNLIRAG 894  
 QY 881 FVSGTGRYHFNDEYELLRCGSIECRPHANNYNINCSKFRF 922  
 DB 895 FSVRAANHFPVNPHEIFGAFPEVRSSSRNNYNLNSKRCF 936

RESULT 9  
 ID 09J542 PRELIMINARY; PRT; 936 AA.

AC 09J542  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).  
 GN PMP 7 OR CP0308.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138.  
 RX MEDLINE-2030349; PubMed-10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Uuchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA".  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE-20150255; PubMed-10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.,  
 RT "Genome sequences of Chlamydia trachomatis Moyn and Chlamydia pneumoniae AR39".  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL; AP002546; BAA98653.1;  
 DR EMBL; AP002193; AAP8165.1;  
 DR TIGR; CP0308;  
 DR InterPro; IPR002637;  
 DR InterPro; IPR003357;  
 DR InterPro; IPR003368;  
 DR Pfam; PF02385; OMP; 1.  
 DR Pfam; PF02415; DUF145; 1.  
 DR ProDom; PD004952; 1.  
 SQ SEQUENCE 936 AA; 100105 MW; 3981DB3C950AF95A CRC64;

Query Match 29.1%; Score 1400.5; DB 2; Length 936;  
 Best Local Similarity 35.9%; Pred. No. 3.4e-83;  
 Matches 338; Conservative 166; Mismatches 391; Indels 47; Gaps 21;

QY 9 PLVFSLLTSLVPTSLSATYISLTPEDSFHGDSQNAERSYVOAGDYVSLTGPVSISSND 68  
 DB 14 PLVFSLLTSLVPTSLSATYISLTPEDSFHGDSQNAERSYVOAGDYVSLTGPVSISSND 69  
 QY 69 --NSALKACFNTSGSVTFAGNHGLEYENNISGTTKEGAVLCCODPQATPARSGFSTL 126  
 DB 70 ALGIPLASGCFLEAGDGLTFQGNQHALKFAFINAGSS-AGTVASTSAADKNLTFNDFSRL 128  
 QY 127 SFIQSGDIKE---QGCIVSKNALMLNNVYFEONOSTKRGALISGAVTVGNDSV 183  
 DB 129 SIISSPSLLSPGQCALSKVGNLSLTGNSQIIFTQNFSSDNGCINTKFFLLSGTSQFA 188  
 QY 184 SFYQNA---TFGGAIHSSGPLOIAVNOAEIFRQONTAKSGSGALYSDDIDIDONAVY 240

```

Db 189 SFSRNOAFTGKGGVYATGTTIENSPIGVSESONLAK-GSGALYSTDNCSITDNOFV 247
OY 241 LPEENALTTAIGKGA VCCLPITGSSSTPVPITGSDNKO LVEERNHNSIMGGA IYARL 300
Db 248 IFGNSNMEAAOAGGAGICCTTDTK-----VTLGNKNLSTNTALYGGAIYGLKV 301
OY 301 SISSGGPTLFINNISYANS-ONLGAIAIDTGEISLSAEKGTITFGNRT---SLPFLN 356
Db 302 SISAGGPTLFQSNISGSSAGGCAININISAGELIASATSGITTFNNNVTGNSSTSTRN 361
OY 357 GHLLONAKFLKLOARNGCSIEFYDPTS--EADGSTOLINDPKNK--EYTGTLFSG 412
Db 362 AIINIDTAKTYSIRATGOSIIFYEDTPTNGTAASDTITLNLINADNSEEYEGAIYFSG 421
OY 413 EKSLAND---PROFKSPIPONVNLISAGYLVIKGAETVSKFQSPSHVLDLGTFLIA 469
Db 422 EKSLPEKAIAANVTSTIROPAVLARDVLKRGVYTFEDLTQSRSRLIMOGTTLSA 481
OY 470 SKEDIAITGLAIDISLSSSTAAYIKANTANKOISVTDSEIELISPTGNAYEDLRMNSQ 529
Db 482 KEANLSLNGLA VNLSSLDGTNKAAL-KTEADAKNLSLSGTALIDTREGSPYENHNLSKAS 540
OY 530 TFFPLSL-EPGAGGSVYTAGDPLV---SPHYGFGQNMKLAN-TGINKVGEFFMDKIN 584
Db 541 TFFPLSLTAGANGTITGLALSTLTQEPETHGYOGNMQLSWANATSSKIGSINMTRTG 600
OY 585 YKRPPEKGNLVNITLGMNAVNRSLMOVETHASSLQTRGLMIDGIGNFHYASSEDN 644
Db 601 YIPSPERKSNLPLNSLMGNIDIRISNOLIETSSGEPFERELMSIANEFPRDSMPT 660
OY 645 IRYRHNSGGVLSVNNELTPKHYSMAFSOLFSDKDYAVSNNEHYMYLGSYLQYTTSL 704
Db 661 HGRHRISGVALGCTTATPREDQLTFAFCQLFARDNRHNTGKNHGDYTGASLTFHHTGL 720
OY 705 GNIFRY----ASNPNVNVGILSRFLONPLMFHFLCAIGATNDMKITYANFPWKNS 760
Db 721 FDIANFLMCKATRAP---VWLSISQIILPLS-FDAKFSYLTDNMHKTYTNTNSIIGS 775
OY 761 WRNCAIECGGSMPLLVFENGRLFOGAIPEMKLOLVYAQDGFKEETADGRFNSGLT 820
Db 776 WRNDACADIGASLP-FVLSVPLLKEVEFFVAYQIYIAHQDFYERYAGRAFNSSELI 834
OY 821 SISVPLGIREFKLALSDVLYDFSFYIPDIFRDKPSCEALYISGDSMLVPAHVSRA 880
Db 835 NVEIPIGVFERBDSKSEKGYDILTYIADAYRNRPCQSLIASANMVAYGTNLARQG 894
OY 881 FVSGGTGRYHFNDETILCRGSEICRPHANAYINCSKRF 922
Db 895 FSVRAANHFOVNPHEIFGQFAFEVRSSSRNYNTNLGSKCF 936

RESULT 10
O9RB65 PRELIMINARY: PRT: 928 AA.
ID 09RB65: 09RB64: 09S6P2; 13, Created)
AC 01-MAY-2000 (Tremblrel, 15, last sequence update)
DT 01-OCT-2000 (Tremblrel, 15, last sequence update)
DE 01-MAR-2001 (Tremblrel, 15, last sequence update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (OUTER MEMBRANE PROTEIN 5
DE PRECURSOR) (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
GN PMP_10 OR OMP5 OR CP0303.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OY NCBI_TaxID=83558;
RN SEQUENCE FROM N.A.
RP STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shira T., Ishikawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishikawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CW029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310;
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birke Lund S.;
RT "Chlamydia proteins containing the GAI-repeat belong to a subfamily
RT of autotransporting pathogenicity factors."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Unayam L.A., Ulfteback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AP002546; BAA98657.1; -
DR EMBL: AJ133034; CAB37071.1; -
DR EMBL: AB002192; AAF38160.1; -
DR TIGR: CP0303; -
DR InterPro: IPR003357; -
DR Pfam: PF02385; OMP; 1.
KW Signal.
SQ SEQUENCE 928 AA; 97229 MW; 0590D5206A1DD0E1 CRC64;

Query Match 28.9%; Score 1393; DB 2; Length 928;
Best Local Similarity 36.1%; Pred. No. 1e-82;
Matches 344; Conservative 162; Mismatches 374; Indels 74; Gaps 27;

OY 10 LVFSLLT--LSVDFSLSATITSLTPEDSFHGSQNAERS-YVQAGDVYSLTGDVYSIN 66
Db 8 LVLSLTACFTSCSYFAAETANIGPDSFGSTNGTTPKNTTGIDTYLLGDTLLQN 67
OY 67 V-DNSALNKAFCNVTSGSVFACNHHGLYFNINSSGTYKGAVALCCODPATARFGSFT 125
Db 68 LGDSALTRKCFSDTIESLFAKRGYSLSFLNKS--SAGGALSVTDR-NLSLGFSS 124
OY 126 LSEFQSPGDI---KEGGLYSKNALMLLNYYVFEQNSKTKGGAISGANTYIVGND 181
Db 125 LTFLAAPSSVITTPSGGAVKCGDILFDNNGTILFKQDYCEENGCAISPKNLSLNKSTG 184
OY 182 SVSFYQNAATF---GGAHSSGPIQIAVQAETIRPQNTARKNGSGGALYSDDDIDIDON 237
Db 185 SISFEGKRSATGKKGALCATGTVDITNTAPTLFSNNIAE-AAGGALNSTGNCITGN 243
OY 238 AYLVFRENALTTAIGKGA VCCLPITGSSSTPVPITGSDNKO LVEERNHNSIMGGA IYA 297
Db 244 TSLVFSNSTATAA-GNGGAL-----SGDAD-----VTISGNQSVTSGNAAVANGA IYA 293
OY 298 RKLSSIS--GGPTLFINNISYANSOINLGAIAIDTGEISLSAEKGTITFGNRT--TSLP 353
Db 294 KRLTLAGSGGGGSGISFNINIVQGTAGNGCAISILAGECSSLAEDGIDFENGNAIYATTP 353
OY 354 FL---NGIHLONAKFLKLOARNGCSIEFYDPTS--ADGSTOLININ-GDPKNK-EYTG 406
Db 354 QYTKRNSIDIGSTAKTITNLNKAISGHISFEYDPTANTAPADSTITLNLNKNADGNSIDYSG 413
OY 407 TIFSGEKLSDPNDP--DEKSTIPQNVNLISAGYLVIKGAETVSKFQSPSHVLDL 463
Db 414 SLVSGEKLSEDAKVAADNLITSLKQPVITLPAQNLVLRKRVITLDTGFTYPAASSYIMDA 473
OY 464 GTKLIAKEDIAITGLAIDISLSSSTAAYIKANTANKOISVTDSEIELISPTGNAYEDL 523
Db 474 GTLLAKSTEEVTLTGLSIPDLSGEK-KVYIAAASAKNVALSGPILLDNGNAYENH 532
OY 524 RMNSOTFPILLSLEPGAAGSVYTAGDPLV---VSPHYGFGQNMKLAN-----TGINKV 575
Db 533 DLGKTQDFSEVOL--SALGATTTDVPAPVETVATPHYGQGTGWTGMDTADTAPEKTKT 590

```



```

RP SEQUENCE FROM N.A.
RC STRAIN-VRI310;
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birkelund S.;
RT "Chlamydia proteins containing the GGA-repeat belong to a subfamily
RT of autotransporting pathogenicity factors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CML029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammell C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CML029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL; AJ01311; CA04672.1; -
DR EMBL; AJ133034; CAB37072.1; -
DR EMBL; AE001628; AAD18593.1; -
DR EMBL; AF002546; BAA98658.1; -
DR InterPro; IPR003357; -
DR InterPro; IPR003368; -
DR Pfam; PF02385; OMP; 1.
DR Pfam; PF02415; DUF145; 1.
KW Signal.
FT CHAIN 1 17 POTENTIAL.
FT CHAIN 18 928 OUTER MEMBRANE PROTEIN 4.
SQ SEQUENCE 928 AA; 98903 MW; 788BCDD62C911402 CRC64;

Query Match 27.9%; Score 1344; DB 2; Length 928;
Best Local Similarity 35.3%; Pred. No. 1.7e-79;
Matches 337; Conservative 152; Mismatches 385; Indels 80; Gaps 22;

QY 10 LVFSLTSLVFDTSLSATTTISLPEDSFHGDSONAERSYVQAG-----DYSLTG 61
DB 14 LAFSCHLOSLANLEL-----LSPDSEFNG-----NIDSGTFPKTSATTSYLTG 58
QY 62 VSTISNV-DNSALNKACFNVTSGSVTFAGNHGLEYENNISGTFKEGAVLCCOPQARF 120
DB 59 VFYFPEKGTPLSDSCFKQTTDNLITFLNGHSLTFPGFDAGTAGAASATTANKNLT--F 116
QY 121 SGRSTLSFIOSPGD--IKEGCCLYSKNAALMLNNVYVRFQNSDKTKGASIGANTYIV 178
DB 117 SGRSTLSFDSPTVYTGGLTSSAGCVLENIRKLVAAGNSTADGAIKASFLTG 176
QY 179 NYNSVSYQN-AATFGAIISSGFLQIAVNOAELRPAQNTAKNGSGALISDGDIDIDON 237
DB 177 TSGDALFSSNNSSSTKGAITTAGARIANTGTGYRFLSNIAST-SGGAIDDEGSIISNN 235
QY 238 AAVLFRENEALTTAIGKGVVCCPTSGSSTPVYTFPSNKOIYERHNSIMGGAIIYA 297
DB 236 KELYFEGNAKKT---GGAICNTKASG---PELITSNKTLIFASNAETSGAIIHA 287
QY 298 KKLSSSGGFLTFI-NNISVANSQNLGAIAIDTGEISLSAEKGTTFPOGN-----RT 350
DB 288 KKLALSSGGFTELRNNVSSATPK--GGAISIDASGELSLSAETGNTTFFRNLITTTGST 345
QY 351 SLPLFNGIHLQAKFLKQARNGCSIEFYDPTSEADSGTQNLINGDPKN--KEYTGTI 408
DB 346 DPKRRAININISGKFTLELAANKHRTFFPYDPTTSEGTSSDVLIKNGSGALNPGGTI 405
QY 409 LFGSGKSLANDPR---DKRTIPQNVLSAGYIVIKGAVNTYSKFTQSGSHVLDLGT 465
DB 406 LFGGTLTTLADELKVADNLKSSFTQVPSLSSGKLLLOKGVLTLESTFSQEGASGLGMDSGT 465

```

```

QY 466 KLIASKEDIAITGLAIDIDSLSSSTAAYIKANTANKQISVTSIELISPTGNAYEDLRM 525
DB 466 TLSTAGSITITNLGINVDSLGLKQPVSLAKGASNKVI-VSGRLNIDIEGNIYESHMF 524
QY 526 RNSCTPFLLEPAGAGSVYTAGDLPY-----SPHYGQGMKWLAV---TGTGNKVG 577
DB 525 SHDPLFLAKITVDADVDVTVDISLLIPPAEDPNSEYGGQGMVNMVTTDTATNKEAT 584
QY 578 FPMDKIYKRPKEGNLVNILLGNANVANRSLMOVQETASSLOTDRLGILGIFNFH 637
DB 585 ATWTKGFVPSPEKSKALVCNTLMGFTDIRSLQQLVEICATGEHKGWVSMTNFIH 644
QY 638 VSASEDNIRYRNHNSGYYLVSNNETPKR-HYTSMAFSQLSRKDYAVASNNEXRMVLG 666
DB 645 KTGDNKKGRFHRTSGGVYIG-GSAHFKDLPFAFCHLEAPKDCFIANNSTRYGGTL 703
QY 697 LYQYTSLSGNIIFYASRNPVNVNGILSRFLONPL-----MIFHELCAYHATNDK 748
DB 704 FFYHSHTL-----QPNYRLRGRKFSESAIEKFPREIPDLADVOVFSHSDRME 754
QY 749 TDYANFPMVKNMNRNMCWATECGSMPLVFEENGRLFOGAIIPFKQLQVAYAGDFRET 808
DB 755 THYTSLSPEGSMSNECTIAGIGIDLPVLSNPHLFTKTIPOKVEAMTVYSONSPRESS 814
QY 809 ADGRFESNGSLTSLVPLGIRFEKLALSDQVLYDFSESYLPDIFRDPSCAALVYSGDS 868
DB 815 SDGRFSGIRLNLISIPVAKFVQGDIGDSYTYDLSGFVSDVYRNPNQSTATLVMSPS 874
QY 869 WLYPAHVSRHAYVSGTGRIYHNDYELLRCRSIEDRPAHRYNINCSKFRF 922
DB 875 WKIRGNLSRQAFLLRGSNNYVYNSCELFHYAMELRGSSRNVNDVGTKLRF 928

RESULT 13
Q9K299 PRELIMINARY; PRT; 949 AA.
ID Q9K299;
AC Q9K299;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE POLYORPHIC MEMBRANE PROTEIN G FAMILY.
GN CP0302.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,
RA Barry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwyn M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AE002192; AAF38159.1; -
DR TIGR; CP0302; -
DR InterPro; IPR003357; -
DR InterPro; IPR003368; -
DR Pfam; PF02385; OMP; 1.
DR Pfam; PF02415; DUF145; 1.
SQ SEQUENCE 949 AA; 101357 MW; A00B09E16C699BE3 CRC64;

```

```

Query Match 27.9%; Score 1344; DB 2; Length 949;
Best Local Similarity 35.3%; Pred. No. 1.7e-79;
Matches 337; Conservative 152; Mismatches 385; Indels 80; Gaps 22;

QY 10 LVFSLTSLVFDTSLSATTTISLPEDSFHGDSONAERSYVQAG-----DYSLTG 61
DB 14 LAFSCHLOSLANLEL-----LSPDSEFNG-----NIDSGTFPKTSATTSYLTG 58

```

QY	410	DYSGSIVSSGKLSDEDAKADNLTSLKQVYTLTAQNLVLRKRYTLDTGFPQTAGSSV	469
Db	412	dysglvifsgklsadeakaadntfsltkpqlaasgltakgnveldivgftqtegstl	471
QY	470	IMDAGTTLKAKTEEWYTLTGSLIPDSSGEGKKVYIAASAKSNVALSGPILLIDNOGNAY	529
Db	472	lmpgqtklkaateaistlklvdlaisalegnkvsietaganckltlslpvlfgdssgnly	531
QY	530	ENHDLGKTODSFYFOLALGATATT-----DPAVVPVATP--THYGQGTWGMVWDDT	582
Db	532	eshlmg----afgqplvftaataasadyldalltspvgtprphhygqghweatwad--	585
QY	583	ASTPKRTATLAWNTGCLLPNPEKQGPLVPNSLMGSSSDIQAIQGVTERALTCLSDRCF	642
Db	586	--tslaksqtlwtvltnpnperrassvpdtslwasfcltdrltqlgmltsqansilygrjl	643
QY	643	WAAVAVAPFLDKDKGKERKTRHKSQGVATIGGAQTGSENLISAFQOLPESDMDFLVAKN	702
Db	644	wasgtanlfhkdksgtlnqatrhksygyvgssaedfsemlfvaicqllykxkdlfiven	703
QY	703	HTDQVAAAFYQIHTCECSGF----IGLDLKLGSMSHKKPLVLEGOLATSHVNSNDLTKTK	758
Db	704	tslhylaslyqlhnaftfgyglmpmsfgsftlml----kdrlllnaqlsystklnmdmty	759
QY	759	TAYPEVKGSMGNNAFNMMLGASSHSY--PEYLHGFDTYAPYIKNTLYTIRODSFSEKGTES	817
Db	760	tsyleaagswmnsgalelgslalylpkeapffqgyflrklqgavysrqmktkeesgaea	819
QY	818	RSFDSNLFNLSLPIGVKFEKFSQDNDFSYDLTSLVYPDLLRNDPKCTALVITSGASWET	877
Db	820	rafddgdgvncslpvgtrtlekisedekmfeislaysgdvyrkmpartslmwsgawts	879
QY	878	YANNLAQALQVRAGSHYASPMKEVVGQVFEVFRGSRITRYNDLGGKRPQF	928
Db	880	lckhlarqafaaagshltslphvelsgseaaaydrgshahynvdcqlrysf	930

RESULT	12
AAW88424	ID
AAW88424	standard; Protein; 930 AA.
XX	
AC	AAW88424;
XX	
DT	26-APR-1999 (first entry)
DE	Chlamydia pneumoniae surface exposed protein Omp11.
XX	
KM	Omp11, outer membrane protein 11; surface exposed protein;
KM	antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.
XX	
OS	Chlamydia pneumoniae.
XX	
PN	WO9858953-A2.
XX	
PD	30-DEC-1998.
XX	
PE	19-JUN-1998; 98WO-DK00266.
XX	
PR	23-JUN-1997; 97DX-0000744.
XX	
PA	(BIRK/) BIRKELUND S.
PA	(CHRI/) CHRISTANSEN G.
XX	
PI	Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI	Mylind P;
XX	
DR	WPI: 1999-105610/09.
DR	N-PSDB: AAX06823.
PT	Species-specific test for identifying mammals infected with
PT	Chlamydia pneumoniae - comprises detecting antibodies specific for
PT	outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT	these proteins

XX Claim 7; Page 63-65; 115pp; English.

XX This polypeptide comprises the novel 97.6 kDa surface exposed

PS protein Omp11 of the human respiratory pathogen Chlamydia

XX pneumoniae. Its amino acid sequence was deduced from DNA (see

CC AA066823) isolated from a C. pneumoniae expression library. The

CC invention provides 12 novel surface exposed proteins, Omp4-Omp15

CC (see AA068411-28), and nucleic acid sequences encoding them (see

CC AA066816-27). A new species specific test is claimed that is used

CC to identify mammals (including humans) infected with Chlamydia

CC pneumoniae. The test comprises detecting antibodies specific for

CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer

CC membrane proteins, especially by PCR. The proteins are also used

CC in the diagnosis of C. pneumoniae infection in mammals. The

CC nucleic acids and proteins can also be used in the immunization of

CC mammals, the nucleic acids being particularly useful as DNA

CC vaccines for effecting in vivo expression of antigens. The

CC vaccines may also prevent atherosclerosis and bronchial asthma,

XX which are possibly associated with C. pneumoniae.

XX

SQ Sequence 930 AA;

Query Match	42.2%	Score 2024:	DB 2024:	length 930:
Best Local Similarity	44.5%	Pred. No. 8,3e-125:		
Matches 423:	Conservative 158:	Mismatches 326:	Indels 44:	Gaps 17/:
OY	1	MKSQSPMLVLSSTLACFTSCSTVFPAATAENIPSPSPDQSTWGTGYTPKNT--TTGIDVT	58	
DB	1	mihplhklilistvlrplllislaatyadaaslsptdsfdgaggs-ftfpkstcdangtvy	59	
OY	59	LTGDTITLMIAGSAAALATKQCFEDTYESLFAKQGLSLPLINKS--SABEALSVTTDKNL	117	
DB	60	lsagnvyindagvgtalgtcgtctettgdltftfgkyysfstntvdagsnagaastadkxl	119	
OY	118	SLTGFSSLTFLAAPSSVITTPSGKCAWCGDLTFDNNCTILFKODYCE--NGCAIST	174	
DB	120	tttfgfnsliaaapt--tvasqkctlsaaqnlntldnglftfsqvnseannngait	177	
OY	175	KMLSLKNNGSLSPFEGKNSSATGKKGGAICACGYDITNNTPLEFSNNIAAAGAIIS	234	
DB	178	krlslsgnsssltfisnsek--klygalysaaaslsqntqylvfmmnkgctsggalgf	234	
OY	235	TGNCITTTGTSVLVFSFENSVTATAPAGAL---SGDA-DVTISGQSVTFSGNAVANG	289	
DB	235	eassstltgnsslffisgntcataadagyggaelyecktgcetpbltllsgnksltifeanssvtgg	294	
OY	290	AIYAKKLTLASGGGGISFSNNIYOGTTAGAGAIISIIAAGECSLSAEAGDITTFNGNALV	349	
DB	295	aicahngldisa--agplifsmrcgntaagkgkgaialadsgslsisaangdtflfgnlt	352	
OY	350	ATTPOTTKRNISIDIGSTAKITMLRAISGHSIFFYDPIRANTAADSTDITLNLKADAGNST	409	
DB	353	stsaqtstnaiaylssakltlnlraagqsylyfyqplasnctq--asdvltlnqpdnspl	411	
OY	410	DVSGSIVSGEKLSDDEAKVADNLSTTKOPATPLAGNLVYLRGTYLDPKGTQPAAGSV	469	
DB	412	dyssglvifsgklsadeakaadnfsllkqplalasgltalkgnvelndngtftqegstl	471	
OY	470	IMDACTTKASTEETLTGLSLPVDLSGEGKKVVAASASAKNVALSGPITLLDNOGNAY	529	
DB	472	lmqptklkadteaalsklvlvdalaalegnksvstetaganktlltclplvrgdssgnfly	531	
OY	530	ENHDIAGKIQDSEFVOLSALGATTT-----DVPAVPIVAAP-THYIGQGTWGMATVDDT	582	
DB	532	eshltng-----atfpvlvftaataasdiydalltspvqdepehygyvgmheatwad--	585	
OY	583	ASTPTKATLALTAMTWTGVLNPERGCPVLPVNSIMSGFSDIOLAGVIERKSLTLCSDSGF	642	
DB	586	--tstaksgtcmwvtltygpnperriassvpdsllwasfdlflrlqqlmsqanslyqgrgl	643	
OY	643	WAAGVANTLDKRRKEKKRRYHSGGYAIGGAAGQCSENLISFACQLEFGSDKDLVAKN	702	



```

Db 644 wasgtanfhkxsgnqatrfhksygiyvgssaedfsenlfsvafcglfqkdkdliven 703
QY 703 HTDTYAGAFYIOHTTECSGF-----IGCLLDKLPQSMHKKPLVLEGOALAYSIVSNDLTKY 758
Db 704 tslnylaslylqhnaflyglpmsfgsftdmf---kdlpdlinaqlsystkndmdtry 759
QY 759 TAYPEVGSWGNNAFMNMLGASSHSY-PEYLHCFDTPYAPYKLNLYIRDDSFSEKGTG 817
Db 760 tsypeagsgwtmnsqalelqgslalylpkeapffgyfplfkfayvsrqnfkesgae 819
QY 818 RSPDDSLNFLSLPIGVKFEKFSDCNDSYDLTSLYVDLIRNDPKCTTALVIGASMET 877
Db 820 rafddqglvncslpvgirlfkesedeknfelslanlgdyvkrpirtsimvasaawts 879
QY 878 YANMLARQALQVRAGSHYAFSPMEVVGQVFEVGRSSRYIVNDLGKFGQ 928
Db 880 lcknlarqafiaasgshlclsphevsgeaayelrsgahynvdcglystf 930

RESULT 13
AAY69368
ID AAY69368 standard; protein; 945 AA.
AC AAY69368;
DE 19-JUN-2000 (first entry)
XX Amino acid sequence of the CPN100394 polypeptide.
XX CPN100394; Chlamydia infection; immune response; vaccine.
XX Chlamydia pneumoniae.
XX
FH Key Location/Qualifiers
FT Peptide 1..43
FT /note= "signal peptide"
PN WO200011183-A2.
PD 02-MAR-2000.
XX
XX 18-AUG-1999; 99WO-IB01449.
XX
XX 20-AUG-1998; 98US-0097187.
XX 20-AUG-1998; 98US-0097188.
XX 20-AUG-1998; 98US-0097189.
XX 20-AUG-1998; 98US-0097190.
XX 20-AUG-1998; 98US-0097195.
XX 20-AUG-1998; 98US-0097196.
XX 20-AUG-1998; 98US-0097197.
XX 27-AUG-1998; 98US-0097191.
XX 17-AUG-1999; 99US-0376770.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP;
PI
XX
DR N-PSDB; AAZ61508.
XX
XX Novel antigens and corresponding DNA molecules that can be used to
XX prevent, treat and diagnose disease caused by Chlamydia infection in
XX mammals, especially humans -
XX
XX Claim 18; Fig 13A-F; 201pp; English.
XX
XX AAY69362-69 represent Chlamydia pneumoniae polypeptides. The
XX polypeptides are present in the bacterial membrane structure, in the
XX external vicinity of the membrane structure, in the inclusion membrane
XX structure, in the external vicinity of the inclusion membrane structure,
XX and in the cytoplasm of the infected cell. The polypeptides may be
XX used to prevent, treat and detect the presence of Chlamydia infection

```

```

CC and/or the presence of Chlamydia in a sample. The polypeptides may
CC also be used to induce an immune response in a mammal. The vaccine
CC vector comprising the polynucleotides is used to induce an immune
CC response in a mammal. Antibodies directed against the polypeptides
CC may also be used therapeutically to treat and/or prevent a Chlamydia
CC infection.
XX
XX Sequence 945 AA;
SO
Query Match 40.4%; Score 1935.5; DB 21; Length 945;
Best Local Similarity 43.3%; Pred. No. 5,0e-119;
Matches 417; Conservative 149; Mismatches 344; Indels 53; Gaps 16;
QY 1 MKSQFNLVLSSTLACTSCSTVPAATENIGPDSFGSTNTGYTPKNT--TGIDYT 58
Db 1 mktsvmlt---allcgasavlyhaatclpnpedgflfgenltfipkstdaagtlts 57
QY 59 LTGDTITLQNLGDSALATKRCFSDTTESLSPAGKGYSLFLNKSACGALSVTTDKNL 118
Db 58 ltgevllydpqkgslyctctfvetagdlftfgngtlklfsvdaglanlavhvgsknls 117
QY 119 LTGSSSLTFLAAPSVTTPBGRKAVKCGDLTFDNNGTILFKDYCEENGALSTKNLS 178
Db 118 ftdflslvitespkrsavlt--gkgslyslgavqlqdlntlvitsnaasvdygvlkgnsc1 175
QY 179 LKNSTGSISEFGNKSATGKGAICATGVDITNNAPTFLFSNNIEAAGAINSTGNC 238
Db 176 lqglnksaifgngts---kkgaisltgltienngltkfenkavtsagaldigaas 232
QY 239 TTIGNTSLVSENSVLTATAGNGALSGDADVT----- 270
Db 233 tftanhellfegnktsqnaangalncsgdlftcdntslllgenstmqdgalestgtls 292
QY 271 ISGNQVTFSGNQAVANGAATYAKKLLASGGGGSFNSNNIVGTAGNGAISTILAAG 330
Db 293 ltsqslvntigntsqgkqgkgaasalkil-99gggajlsmvvhacp-lygaalfntcg 350
QY 331 ECSLSAAGDITTFNGNAIVATTPO-TTKRNSIDIGSTAKITNLALISGHSITFFDPTAN 389
Db 351 slqftfggdvifegngvtttapatckkrvnhlesctakwtglaasgnalfydptlt 410
QY 390 TAASTDPTLNLKADAGNSTDYSGSIVFSGEKLSDEDAKVAADNTSTLKOPVTTAGNLV 449
Db 411 dtg-asdnrlrinesangk--lsgsivfsgerlstaesa-laenltslngpvtlvegsly 466
QY 450 LKRGVLTDTGFGTQTAGSSVYMDAGTTLKASTEBEVTLNGISIPVDSIGECKKVIAASAA 509
Db 467 lkqgvclltqfsgqepescllldgtclkaastedvltlnslmadcllygknplivasaa 526
QY 510 SKNVALSGPILLDNOGNAYENHDLGKTODFSFVQLS--ALGTATTTDVAPPTVATPT- 566
Db 527 nkoitlgtlalnadaqafeyhncilqsgdysfvlkpsgaagglittdasgkplevapsr 586
QY 567 -HYIGGTGWTWVDTPASTPKTKTATLANTNGYLLNPEKQGLVNSLMSGSSDIOAI 625
Db 587 phygvgghwnvqvlypigtcp--sqanlewrtlylpnpergslvlpnslwsgsvdgral 644
QY 626 OGVTERSAITLCSDRGFAGVAFELDKDKGKKRKYRHSKSGAICAQOTCESENLSF 685
Db 645 qelmnassqllqgergywagglanflnrcklne-hgytrhsqvglylvagthafsdactina 703
QY 686 AFCLFSGSDEFLVAKNHTTYAGAFYIOHTTECSGFTGCLLDLPQSMHKKPLVLEGOAL 745
Db 764 syhrnmndmktkytlypeagsgwandvfgylefattyynostflfdyspflrlgctya 823
QY 806 RQDSFSKGTGEGRFSFDSNLFNLSLPIGVKFEKFSDCNDSYDLTSLYVDLIRNDPKCT 865
Db 824 hgedfketsgevrntfsgdflnlaivpvgvkferfsdckrsgsyeltlayvpdvirkdkb-s 882

```

Oy	866	TALVTSIGSAWERYANNLAQAOVRAGSHYATSPMEVEVGQFVFEEVGRSGRIYNDLGK	925
		:     :   :     :   :     :   :	
Dd	883	tatlsgawtschgnmlsrqqlrlgncplmpgievfrshgaletgrssrmynl9gk	942
Oy	926	EQG 928	
	:::	:	
Dd	943	yrf 945	
	RESULT	14	
ID	AAM88423		
XX	AAM88423 standard; Protein; 928 AA.		
AC	AAM88423;		
DX			
DT	26-APR-1999 (first entry)		
XX			
DE	Chlamydia pneumoniae surface exposed protein Omp10.		
XX			
KM	Omp10; outer membrane protein 10; surface exposed protein; antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.		
OS	Chlamydia pneumoniae.		
PN	MO9858953-AZ.		
XX			
PD	30-DEC-1998.		
XX			
PF	19-JUN-1998; 98WO-DK00266.		
PR	23-JUN-1997; 97DK-0000744.		
XX			
PA	(BIRK/) BIRKELUND S. (CHRT/) CHRISTIANSEN G.		
Birkelund S,	Christiansen G,	Knudsen K,	Madsen A;
PI			
PI	Myglind P;		
DR			
N-PSDB:	AAK06822.		
MPJ:	1999-105610/09.		
Species-specific test for identifying mammals infected with Chlamydia pneumoniae - comprises detecting antibodies specific for outer membrane proteins of C. pneumoniae or nucleic acids encoding these proteins			
Claim 7; Page 60-62; 115pp; English.			
This polypeptide comprises the novel 98.4 kDa surface exposed protein Omp10 of the human respiratory pathogen Chlamydia pneumoniae. Its amino acid sequence was deduced from DNA (see AAK06822) isolated from a C. pneumoniae expression library. The Invention provides 12 novel surface exposed proteins, Omp4-Omp15 (see AAM88417-28), and nucleic acid sequences encoding them (see AAK06816-27). A new species specific test is claimed that is used to identify mammals (including humans) infected with Chlamydia pneumoniae. The test comprises detecting antibodies specific for Omp4-Omp15 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used in the diagnosis of C. pneumoniae infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with C. pneumoniae.			
Sequence	928 AA;		

Query Match 40.3%; Score 1931; DB 20; Length 928;  
Best Local Similarity 44.7%; Pred. No. 1.1e-118;  
Matches 423; Conservative 149; Mismatches 338; Indels 36; Gaps 18;

QY	1	MKSQSWLVSSTLACFTSCS--TFPAABE-NIGPSDSFOSSTMTGTYTTRKNTTGTGD--	56
Db	1	mksqshwlvssslalplshlfsfaavehnlplnpsfsg--pdytppagctlnadt	57
QY	57	-YTLTGDTTLQMLGSAALTRKGFCEDTTESISFPAKGKSLSEFLNINKSSAECAALSVTDK	115
Db	58	lynlvgdvsichnagaprtalacsfletctgnlsfghyqfllqnlhdagancctfntcaank	116
QY	116	NLSLTGFSSLTFLPAASSVITTPSGKAVKCGGLTFEDNNGTILFKODYCEENGALSTK	175
Db	118	llsfefaylsjltgtnat---lqgalkstgacslqmsyfcyfgqnfundngalqgs	173
QY	176	NLSLKNSTGSGISFEENKSSATGKKGALCAICGYDITNTNTPLEFSNNINAEACGAINST	235
Db	174	sisl-slnpultfankat---qkqyalysfvgllntlnlnsasifseantaanqalyte	225
QY	236	GNCTITGNTSLVFSNSVYATATAGNGAL---SGAD--VTISGNQSVTFSGNQAVANG	289
Db	230	asfslsnkalsflninstacsaagyalycstapkpvtllnsngelnfngntalsg	285
QY	290	AIYAKKLTLASGGGGGGSFSNNIVQGTAGNGAIIILAGECSLSAEADITFNGNAIV	349
Db	290	aiydnlylss--ggpultfknsaidtaaprgyaladsgslalsalgdlftegtlv	347
QY	350	--ATTPOTTKRNSTDICT--AKITNLRAISHSIFFPDPTANAASTDTLNANKDAG	406
Db	348	kyassqglltnlnslngltlnaklvqlrsqgnltlydplntlnlaasaldalnogpla	407
QY	407	NSTDVSGSIVSGEKLSEDEKAVDNLTSLTKQVNTTLAGNLVLRKAVTLDTKGFTTAG	466
Db	408	gnpayqglivfsekgeliseaeaaadnlkstlqgpltlagglslsksgvtlvaksfsgspg	467
QY	467	SSVIMDACTTLKASTEEVYTLGLSLIPDLSCEGKKVVYIASASKNVALSGPILLDNOG	526
Db	468	stllmdagcttle-fadgltltnlmlnvdsiketkalkatqasqvtvlsgslsvbpqg	526
QY	527	NAYENHDKGKNODSEFYQSLALGTAT--TPDVPAPVPYAPPTNHYGQGTGCMGTWDDTAS	584
Db	527	nveydevsnmpvfvsciltctadoprnlhtctdaadpleknpinhwygnavalsqedatc	585
QY	585	TPKTKTATLANTCGYLRPNRQGEPLVPNSLMGSPDIOALQGVIERSAALTLCSDRGFWA	644
Db	587	--kskaatlwtlckbgnpnperrgltvaantlwgsvfdvrsldqvlvackvqsgctgrlwc	644
QY	645	AGVANFLDKDKKGEKRRKRRHNSGGVYALGGAQOTSENLISAPQLRGSQDKPLVAKNHT	704
Db	645	eglsinfikdstcklnkgrftrhsaayvavattclashdlitaeafqlygkdrdhlnkra	704
QY	705	DTYAGAFYIQHITECSEFGICLLKLCPGSMWRKRLVLEGQALSHVSNENDLTKTAYAPEV	764
Db	705	sayaaslnhqlatcsls--pslllylpgseseqvlfdaaqtsylyskntmklyttqapkg	763
QY	765	KGSWGNNAFNMMLGAS--SHSYPEYLHCFTDYAPYIKLNTLYIRQDSFSEKGE-GRSFD	822
Db	763	esswyndgcalleasslphtalshegflfhayfllkfwasyibhdsfkekrntllvrsfds	822
QY	823	SNLPLSLPIGVKREKPSDCNDFSYDULTSVYBPLILINDPKRCTTALVYISASWMTYANNL	882
Db	823	gdllnvsvpdlfterfiserneasyeavtlyvadvaytknpctalllnltswkvtgtnl	882
QY	883	ARQALQVRAGSHYAFSPMFEVLGQFVEVRSSSRIRVYNDVLGKRFQF 928	
Db	883	srgagigtragflfyafsnpletvtslnsmelrsgsrsvynadlgykrf 928	

RESULT	15
AAV90239	
ID	AAV90239 standard; Protein; 928 AA
XX	
AC	
AAV90239;	
XX	
DT	29-AUG-2000 (first entry)



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: October 2, 2001, 03:27:12 ; Search time 52.09 Seconds

(without alignments)  
366.824 Million cell updates/sec

Title: US-09-446-677B-4  
Perfect score: 4795  
Sequence: 1 MKSQFSLVLSSTLACTSC.....FEVGRSSRIYNDLGKRFQ 928

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA\*

1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/6C.COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/6D.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No	Score	Query Match	Length	ID	Description
1	291	6.1	1026	1 US-08-194-290-7	Sequence 7, Appl
2	286.5	6.0	1600	2 US-08-617-697-10	Sequence 10, Appl
3	285.5	6.0	1529	2 US-08-728-470-10	Sequence 10, Appl
4	285.5	6.0	1529	4 US-08-719-641-10	Sequence 10, Appl
5	284.5	5.9	1338	2 US-08-728-470-9	Sequence 9, Appl
6	284.5	5.9	1338	4 US-08-719-641-9	Sequence 9, Appl
7	284.5	5.9	1599	2 US-08-617-697-9	Sequence 9, Appl
8	284	5.9	1026	2 US-08-614-377A-7	Sequence 7, Appl
9	284	5.9	1026	4 US-09-142-648B-7	Sequence 7, Appl
10	259	5.4	1536	1 US-08-538-682-2	Sequence 2, Appl
11	259	5.4	1536	1 US-08-302-832-2	Sequence 2, Appl
12	259	5.4	1536	2 US-08-530-198-2	Sequence 2, Appl
13	259	5.4	1536	2 US-08-469-880-2	Sequence 2, Appl
14	259	5.4	1536	2 US-08-728-470-2	Sequence 2, Appl
15	259	5.4	1536	2 US-08-617-697-2	Sequence 2, Appl
16	259	5.4	1536	4 US-08-719-641-2	Sequence 2, Appl
17	248.5	5.2	1612	1 US-08-169-927-2	Sequence 2, Appl
18	227	4.7	1477	1 US-08-038-682-4	Sequence 4, Appl
19	227	4.7	1477	1 US-08-302-832-4	Sequence 4, Appl
20	227	4.7	1477	2 US-08-530-198-4	Sequence 4, Appl
21	227	4.7	1477	2 US-08-469-880-4	Sequence 4, Appl
22	227	4.7	1477	2 US-08-728-470-4	Sequence 4, Appl
23	227	4.7	1477	2 US-08-719-641-4	Sequence 4, Appl
24	227	4.7	1477	4 US-08-808-599A-24	Sequence 24, Appl
25	222.5	4.6	1160	3 US-08-348-353-17	Sequence 17, Appl
26	215	4.5	1248	2 US-08-465-965-17	Sequence 17, Appl
27	215	4.5	1248	2 US-08-465-965-17	Sequence 17, Appl

28	215	4.5	1248	3 US-08-465-966-17	Sequence 17, Appl
29	211.5	4.4	2353	4 US-09-377-155-33	Sequence 33, Appl
30	211.5	4.4	2353	4 US-08-913-942-4	Sequence 4, Appl
31	209.5	4.4	1912	1 US-08-409-995-4	Sequence 4, Appl
32	209.5	4.4	1912	1 US-08-685-467-4	Sequence 4, Appl
33	207	4.3	2123	4 US-08-968-685A-10	Sequence 10, Appl
34	194.5	4.1	674	1 US-08-317-522A-3	Sequence 3, Appl
35	194.5	4.1	674	1 US-08-439-818A-3	Sequence 3, Appl
36	194.5	4.1	674	2 US-08-731-965-3	Sequence 3, Appl
37	194.5	4.1	674	2 US-08-738-975-3	Sequence 3, Appl
38	194.5	4.1	674	2 US-08-728-626-3	Sequence 3, Appl
39	194.5	4.1	674	3 US-08-808-599A-3	Sequence 3, Appl
40	193.5	4.0	749	1 US-08-317-522A-2	Sequence 2, Appl
41	193.5	4.0	749	1 US-08-439-818A-2	Sequence 2, Appl
42	193.5	4.0	749	2 US-08-731-965-2	Sequence 2, Appl
43	193.5	4.0	749	2 US-08-738-975-2	Sequence 2, Appl
44	193.5	4.0	749	2 US-08-728-626-2	Sequence 2, Appl
45	193.5	4.0	749	3 US-08-808-599A-2	Sequence 2, Appl

## ALIGNMENTS

```

RESULT 1
US-08-194-290-7
Sequence 7, Application US/08194290
Patent No. 5500353
GENERAL INFORMATION:
APPLICANT: Smith, John
APPLICANT: Bingle, Wade H
TITLE OF INVENTION: Bacterial surface protein expression
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Shlesinger, Arkwright & Garvey
STREET: 3000 South Eads Street
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194,290
FILING DATE: 09-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Garvey, George A
REGISTRATION NUMBER: 17737
REFERENCE/DOCKET NUMBER: 5946-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-684-5600
TELEFAX: 703-836-5288
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1026 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-194-290-7

Query Match 61% Score 291; DB 1; Length 1026;
Best Local Similarity 24.1% Pred. No. 1.4e-15;
Matches 191; Conservative 97; Mismatches 286; Indels 220; Gaps 40;
QY 33 PDSFDGST---NTGTY---PKNTTGIDYTLTGDTLQNGDSALTRGCFSDPTESL 86
DB 243 PSSGVSGSLSLTGTDTLTLGTANNDFVAGEVAGATL--TVGDT--LSCGAGTDLVANN 299
OY 87 SPAG-----KQSLSLFNIRKSSAGCALSVTTDKNLSTNGFSSL--TFLAAPSIVITPPSG 140

```

Tue Oct 9 10:45:12 2001

us-09-446-677b-4.rai

Page 2

Db 300 QAAAYALPTGVITISCIEMNTVSGAAILNTLSS--GVGLALMTNTSGAAYTAGAG 357  
QY 141 KG-----AVKGGDLEFNNNGTILFKODYCENGCALSTKLNKSTGSGIS 187  
Db 358 QNLTATTAQAANNVAADGAGANTVASTGV-----TSGTITVGNANSAAGTGVAS 407  
QY 188 FEGKSSATGKGGALCATGTVTNTTAPLESNNIAEAGAINST---GNCITITN- 243  
Db 408 V-ANSSITT-----TGAIATVGTAVT-----VAOTAGNAVNTLTQADVTVTGS 452  
QY 244 -TSVSESNVATAGN--GGAISGADVTISGNSQVTSQNOAVANGAIVAKLLTAS 300  
Db 453 STTAVTVQTAATAAGATVAGRVGAVTITDSAAASKTAKTAT-----VTLGS 502  
QY 301 GGGGISTSNIVGTTAGNGCAISILAGECSISAEAGDITFNGNAIVATTPOTTKRNS 360  
Db 503 FGAATID--SSALTTVNSLSTGTSIGI--GRGLATATPTANTLTINVLNGLTGTGATDSEAA 560  
QY 361 IDIGSTAKITNLRAISGHISFEYDPTANTAADSTDLNKNKADAGNSTDYSGSIFSGE 420  
Db 561 ADDGTT--INAGSTASS-----TASLVADAT--PLNI-----SGD 595  
QY 421 KUSEDAKAVADNLTSTKQPTLTLAGNLVKRGVLTDTK-----GFTQAGSSVIMDAGT 475  
Db 596 A-----RVITTSHTAALITGITVNSVCAITLGAELATGLVFTGGAGADSLILGAT 645  
QY 476 T-----LKASTEVTTLGLSIPVDSL--GEGKKVIA--ASAANKVVALSGPILLIDNO 525  
Db 646 TKATVAGADDTVYTSANLTLAGSGSVNGDGDVLVANVNSSSEADPAFEGFETLRVAG 705  
QY 526 GNAVENHDLKGTQDSEFVLSALGTAT-----IDVAVPVTATPTXG----- 569  
Db 706 AAQSGSHN--ANGFTALQIGATAGATPTNVAVNVGLTIVLAAPGTITVTLANATGSD 762  
QY 570 -----YQGTWGMWVD-----DTASTPKETATLA----- 594  
Db 763 VFNLTSSSALAALAGVATVETVNIATDINTAHVTLTLQATSAKSIYVTGNAGLN 822  
QY 595 WNTGYLPNEROGPLVPSNLGMSFSDIQAIOGVIERLSALITCSDRGFAAGVANFLDK 654  
Db 823 LTNNG-----NTAVTSF--DASAVTGT--GSAVTFVS-----AN---T 853  
QY 655 KKEKKRYRHKSGGYAIGAAQTCSENLISFACQLEGSQKDFLVAKNHTDYAGA--- 710  
Db 854 TVGEVVTIRGAGADSLTGS--TANDTIIG-----GAGADLVYTTGGDITFTGGTGAD 905  
QY 711 -FYIHITECSGFI 723  
Db 906 IFDINAIGSTAFV 919

RESULT 2

US-08-617-697-10  
Sequence 10, Application US/08617697

Patent No. 5977336

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/617,697  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 05-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-557  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SRO ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1600 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-617-697-10

Query Match 6.0%; Score 286.5; DB 2; Length 1600;  
Best Local Similarity 22.5%; Pred. No. 6.0e-15;  
Matches 150; Conservative 106; Mismatches 220; Indels 191; Gaps 31;

QY 30 NIGSPSPGGSNN-----TGATPKNTT-----GIDVYLT----- 60  
Db 903 SIAEDSTFKEGASDNLTITGTTNNGTANINIKGVYKLGIDINNKGINITTNASGTOKTI 962  
QY 61 -----GDITLONL--GDSALATKGCPSDTTESLSFAGKGYSL--SELNK----- 101  
Db 963 INGNITNEKGDINIKKADAEIOGINISOKEGNLTSSDKVNTTNOITTKAGVBGGRS 1022  
QY 102 --SSAGCALSVTTDK-----NLSLTGFSLLFLAAPSIVTTPSGKAVKCGGDLTFD- 153  
Db 1023 DSEAEANANLTOTKELKLAGDLNISGFNKALTTAKNSGDLTIGASGNDADAKVFFDK 1082  
QY 154 -----NNGTILFKODYCEENGALISTKNSLKNSTGSG--ISFE 189  
Db 1083 VKDSKISTDGHNTVNLSEVYTSNGSSNAGND--NSTGLTISAKDVTNNVNTSHTKINIS 1140  
QY 190 GNKSSATGKKGALICA--TGTVDITNTTAPLPSNNIAEAGAAGAIN---STGNCITIGNTS 245  
Db 1141 AAAGNVTTKEGTTINATGSEVY-----AONGTIKGNITSONVYATAREN 1186  
QY 246 LVFSENSY-TATAG-----NGG--ALSGADVTISGN---QSVTFSGNOAVAN 287  
Db 1187 LVTTENAVINATSGTVNISTKGTGDIKIGISTSGNVAINTASGNTLKVSNITGGDVTATD 1246  
QY 288 GGAI--YAKKLTLLASGGGGSIFSNNTVOGTTAGNGCAISILAG-----ECSL 334  
Db 1247 ACALTTTGTAGSITISATGTANNTTGTGDIKNGKESSGCVTLATGATLAVANGISGNTVIT 1306  
QY 335 SAEAGDITFN--GNAIVATTPOTTKRNSIDISTAKITNLRAISGHISFEYDPTANT--- 390  
Db 1307 TADSGKLTSTVGSTINGNSVTSOSGDIGT-----ISGNTV---NVTASTGDL 1354  
QY 391 -----ADSTDLNKNKADAGNSTDYSGSIFSGEKLSEDAKVAADNLTSTLKOPY 441  
Db 1355 TIGNSAKVAKKGAATLT--AESGKLTTOGSSITS-----SNGOTTPLAKD 1399  
QY 442 TLTAGNLVLKRGVTLDTKGTOTAGSSVIMDAGTLLKASTEVEVLTGLST-----PVD 494  
Db 1400 SSTAGN--INANVTLTGTGLTGTGDSKINATSGTLINAKDKIDGAASGDRTVNATN 1458  
QY 495 SLGEGKKVYLAASAKNV-----ALSGPILLIDNOGNAYENHDLKGTQDSEFVLSALG 549

Db 1459 ASGSGN--VTAKTSSVNITGDLNTINGNLITISENGRNTVRLR--GKEIDVKYIQ--PG 1511  
QY 550 TATTTDV 556  
Db 1512 VASVEEV 1518

RESULT 3  
US-08-728-470-10  
; Sequence 10, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Matlare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; City: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728,470  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstesser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-633  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0810  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1529 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-728-470-10

Query Match 6.0%; Score 285.5; DB 2; Length 1529;  
Best local Similarity 22.4%; Pred. No. 7.5e-15;  
Matches 150; Conservative 106; Mismatches 220; Indels 193; Gaps 31;

QY 30 NIGRSDSFDSTN-----TGTYTPKNTT-----GIDYTLT----- 60  
Db 831 SIADSTFKGEASDNINITGTFTNGTANINIKGVYKLOGDINNKGGLNITTNASGTOK 890  
QY 61 -----GDTTQONTL-GDSALATKGCFSPTTESLSFPAKGYSL-SPLNIR----- 101  
Db 891 TITNGNTNTEKGLDNKINRADAIEIOIGNISOKEGMLTSSDKVNITTTNOITTKAGVEGG 950  
QY 102 -----SSAEGALSVYTDK-----NLSITGFSSLTFLAAPSSVYITTPSGKAVKCGGLTF 152  
Db 951 RSDSSEAEANLITQTFKELKLAGDLNINISGNKAEITPAKNGSDLTIGNAGSGNDAAKVTF 1010

QY 153 D-----NNGTILFKODYCEENGALSTKNLSLKNSTGS---IS 187  
Db 1011 DKVDSKISTDGHVNTLNSEVKTNGSSNAGND--NSTGLTISAKDVVNNNTSHKTIIN 1068  
QY 188 FEGNKSATGCKKGAICA-TGTVDTNTNTAPLFSNNIAEAGAIN---STGNCTIGIN 243  
Db 1069 ISAAAGNVTTEGTTINATTSVEVT-----AONGTIGNITSONVTAT 1114  
QY 244 TSLVSENSV-TATAG-----NGG--ALSGADVTISGN---OSVTFSGNOAV 285  
Db 1115 ENLVTTENAVINATSGVYNISTKTGDIKGISTSGNVTINATSGNTLKVSNITGQDVVT 1174  
QY 286 ANGAI--YAKKLLASGGGGISFSSNNIVOGTTAGNGAISILAG-----EC 332  
Db 1175 ADAGALTTTAGSTISATGTNNITTKGTGDIKGKVESSSGSVTLVATGTLAVNIGNTV 1234  
QY 333 SLSAEADITPEN-GNAIVATTPOTTKRNSIDIGSTAKTTNRAISGHSIFFYDPTTANT- 390  
Db 1235 TITADSKLFTSVGSTINGTMSVTTSSGSDIEGT-----ISGNIV-----NTASTG 1282  
QY 391 -----ADSTDTLNINKADAGNSTDYSGSIYFSGEKLSEDEAKVADNLTSTLKQ 439  
Db 1283 DLITGNSAKVYAKKNGAATLT--AESGKLTTQTGSSITS-----SNGQTLTLA 1327  
QY 440 PVYTLFAGNLVLRGVTLDTKGFQTAGSSVIMDACTTLKASTEVTTLGLSI-----P 492  
Db 1328 KDSSIAON-INAANVTNTTGTTLTGTGSKINATSGTLTINAKKAKLDGASGRTVYNA 1386  
QY 493 VDSLGECKKVYIAASASKNV-----ALSGPILLDNQGNAYENHDKTQDFSVOLSA 547  
Db 1387 TNAGSGN--VTAKTSSVNITGDLNTINGNLITISENGRNTVRLR--GKEIDVKYIQ-- 1439  
QY 548 LGTATTTDV 556  
Db 1440 PGVASVEEV 1448

RESULT 4  
US-08-719-641-10  
; Sequence 10, Application US/08719641  
; Patent No. 6218141  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Matlare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; City: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/719,641  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992

ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-625  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1529 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-719-641-10

Query Match 6.0%; Score 285.5; DB 4; Length 1529;  
Best Local Similarity 22.4%; Pred. No. 7.5e-15;  
Matches 150; Conservative 106; Mismatches 220; Indels 193; Gaps 31;

30 NIGPSDSFDSSTN-----TGTTPPKNTT-----GIDYTL----- 60  
831 SIADSTFKGEADSNLITGTFTNGTANINIKGVYKIOGDINNKGGLITTNASGTOK 890  
61 -----GDTLQNL-GDSALTKGCFSDTBSLSFAGKGYSL-SFLNIK----- 101  
891 TIINGNTNKGDLNKNIMADAEIQIGNISOKEGMLTISDKVNITNGITIKAGVEGG 950  
102 -----SSAEGALSVYTDK-----NLSITGFSLSFLAAPSIVITTPSGKAVKCGDLTF 152  
951 RSDSSEAEANMLTIQTEKELAGDLNLSGPNKAEITAKNGSDLTIGNASGMDAKKVTF 1010  
153 D-----NNGTILFKODYCEENGGAISPKNLKLNKSTGS-----IS 187  
1011 DKYKDKISTDGHVNTLNSEVKTSSNGSSNAGND--NSTGLTISKADVTYVANNVTSHTIN 1068  
188 FEGNKSATGKKGGAICA-TGYVDITNNTPATLFSNNIAEAGAIN--STGNCITTN 243  
1069 ISAAAGNVTTEKGTINATGVSVEVT-----AQNGTIGKNTISQNVTVAT 1114  
244 TSLVFSNSV-TATAG-----NGG--ALSGDADVTISGN-----QSVFESGNAV 285  
1115 ENLVTEENAVINATSGVNISTRGTGDIKGIESTSGVNNITASGNTLKVSNTIGQDVTVT 1174  
286 ANGCAI--YAKKLTILASGGGGISFSNNIVOGTTAGNGAISILAA-----EC 332  
1175 ADAGALTTTGTAGTISATGTAANITTTKGTINDNGKVESSGCVTLVATGATLAVNGISGTV 1234  
333 SLSAEGADITFN-GNAIVATTPPTTKRNSIDIGSTAKITNLRALISHSIFFEPTITANT- 390  
1235 TITADSGKLTSTVSGTINSVTSSQSDIEGT-----ISGNTV-----NVTASTG 1282  
391 -----AADSTTILNLRKADAGNSTDYSGSIYFSGEKLSEDAKADNLTSLKQ 439  
1283 DLITGNSAKVAKNGAATLT--AESGKLTGTGSSITTS-----SMQDTTLTA 1327  
440 PVTLTAGNLVLRKGVLTDTGFTQTAGSSYVIMAGTILKASTEEVTLTGLSI-----P 492  
1328 KDSIIAGN-INAANVTILNTIGTLTTGDSKINATSGTLTINAKADKADGAASGRTVYNA 1386  
493 VDSLGEKRVVIAASAASKAV-----ALSGPILLDNOGNAVENHDLKQDPSFVOLSA 547  
1387 TNASGSGN--VTAKTSSSVNITGDLNLTINGLNIISENGRVTVRLR--CKEIDVXYIO--- 1439  
548 LGTATTTDV 556  
1440 PGVAVSEEV 1448

RESULT 5  
US-08-728-470-9  
Sequence 9, Application US/08728470  
Patent No. 5928651

GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,470  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1338 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-728-470-9

Query Match 5.9%; Score 284.5; DB 2; Length 1338;  
Best Local Similarity 22.5%; Pred. No. 7.3e-15;  
Matches 151; Conservative 111; Mismatches 223; Indels 187; Gaps 31;

12 STLACFTSCSTVFPAATNENIGPSPD-----SFDGNTNGTTPKNTTGIDYTL----- 59  
646 ANIQATINTYFNWAGSDNNGASNISIAKGAAPKIDINNTSSL--NITNSDTTITRTII 702  
60 -----TGDTLQNL-----LGDNALTKGCFSDTBSLSFAGKGYSL-SFLNIK--- 101  
703 KGNISNKGDLNITDKSDAEIQIGNISOKEGMLTISDKVNITNG-----ITIKAGV 756  
102 -----SSAEGALSVYTDK-----NLSITGFSLSFLAAPSIVITTPSGKAVKCGDLTF 149  
757 EGGRSDESSEANMLTIQTEKELAGDLNLSGPNKAEITAKNGSDLTIGNASGMDAKK 816  
150 LTFD-----NNGTILFKODYCEENGGAISPKNLKLNKSTGS--- 185  
817 VTFDVKYDKISTDGHVNTLNSEVKTSSNGSSNAGND--NSTGLTISKADVTYVANNVTSK 874  
186 -ISFEGNKSATGKKGGAICA-TGYVDITNNTPATLFSNNIAEAGAIN--STGNCIT 240  
875 TINISAAAGNVTTEKGTINATGVSVEVT-----AQNGTIGKNTISQNVTVAT 920  
241 TGNISLVFSNSV-TATAG-----NGG--ALSGDADVTISGN-----QSVFESGN 282



```

Db 921 TATENLYTTEAVINAVINAGSVINISTKTGDIKGIESTSGVNVITAGNTLKVSNITGQDV 980
QY 283 QAVANGAI--YAKKLTLASGGGGISFSNNIYOGTATAGNGAISILAAG----- 330
Db 981 TWTADAGALTTTASGTTISATGNGANNITTKGTDINGKVESSGSVTLVATGATLAVGNISG 1040
QY 331 -ECLISAEAGDITFN--GNAIVATTPOTTKRNSIDIGSTAKITNLRAISGHSIFFPDPTA 388
Db 1041 NTVTITADSGKLITVSTGTINGTNSVTTSSQSDIEGT-----ISGNTV---NNTA 1088
QY 389 NT-----AADSTDLNLNKADAGNSTDYSGSIYFSGEKLSEDEAKVADNLST 436
Db 1089 STGDLTIGNSAKVAKNGAATLT--AESGKLTTQTSSITS-----SNGQTT 1133
QY 437 LKQPVTLTAGNLVLRKGVTLDTKGFOTAGSSVIMDAGTTLKASTEYVTLGLSI----- 491
Db 1134 LTAKDSSIAGN--INAAVTLTNTTGTTLTTGDSKINATSGTLTINAKAKLDGAASGDRV 1192
QY 492 --PVDSIGEGKKVVIASASAKNV-----ALSGPILLDNOGNAYENHDLGKTODFSFVQ 544
Db 1193 VNATNAGSGGN--VTAKTSSVNITGDLNTINGLNIISENGRNVRLR--GKEIDVRYIQ 1248
QY 545 LSAAGTATTTDV 556
Db 1249 ---PGVASVEEV 1257

RESULT 6
US-08-719-641-9
; Sequence 9, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1338 amino acids
; TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-9

Query Match
Best Local Similarity 22.5%; Pred. No. 7,36-15;
Matches 151; Conservative 111; Mismatches 223; Indels 187; Gaps 31;

QY 12 STTACFTSCSTVFATAENIGPSD-----SFDGNTNGTTPPKMTTGIDYTL----- 59
Db 646 ANIQATNTYFENVAGSFDNNGASNIJARAGAKFKDINNTSSL---NITNSTDYTRRII 702
QY 60 -----TGDTLTON-----LGDNALTKGCPSPDTESLSPAGKYSLSPLNIK---- 101
Db 703 KGNISNKSQDNIITDKSDAEIOIGNISQKEGMLTJSSDKVNTNQ-----TTIKGV 756
QY 102 -----SSAGGALSYTTDK-----NLSTGFSSLFLAPSSVITTPSGKAVKCGGD 149
Db 757 EGGKSDSSEABENAMLTITQKELAGDLNLSGFKAETLAKNGSDDLITGNASGNGADAKK 816
QY 150 LTFD-----NNGTLFRQDYCEENGAISTRNLILKNSTGS-- 185
Db 817 VTFPKVAKDSKISTDGHVNTLNSSEVKTSSNGAGND--NSTGLTISAKDVTYNNNTVSHK 874
QY 186 -ISFEGKSSATGKKGAIKA--TCTVDITNTAPTLESNNIAEAGGAIN---STGACTI 240
Db 875 TINISAAAGVYTKEGTINATTSVEVT-----AONGTIKGNITISQVNTV 920
QY 241 TGNLSVSENSV--TATAG-----NGG--ALSGDADVTISGN---QSVTFESGN 282
Db 921 TATENLYTTEAVINAVINAGSVINISTKTGDIKGIESTSGVNVITAGNTLKVSNITGQDV 980
QY 283 QAVANGAI--YAKKLTLASGGGGISFSNNIYOGTATAGNGAISILAAG----- 330
Db 981 TWTADAGALTTTASGTTISATGNGANNITTKGTDINGKVESSGSVTLVATGATLAVGNISG 1040
QY 331 -ECLISAEAGDITFN--GNAIVATTPOTTKRNSIDIGSTAKITNLRAISGHSIFFPDPTA 388
Db 1041 NTVTITADSGKLITVSTGTINGTNSVTTSSQSDIEGT-----ISGNTV---NNTA 1088
QY 389 NT-----AADSTDLNLNKADAGNSTDYSGSIYFSGEKLSEDEAKVADNLST 436
Db 1089 STGDLTIGNSAKVAKNGAATLT--AESGKLTTQTSSITS-----SNGQTT 1133
QY 437 LKQPVTLTAGNLVLRKGVTLDTKGFOTAGSSVIMDAGTTLKASTEYVTLGLSI----- 491
Db 1134 LTAKDSSIAGN--INAAVTLTNTTGTTLTTGDSKINATSGTLTINAKAKLDGAASGDRV 1192
QY 492 --PVDSIGEGKKVVIASASAKNV-----ALSGPILLDNOGNAYENHDLGKTODFSFVQ 544
Db 1193 VNATNAGSGGN--VTAKTSSVNITGDLNTINGLNIISENGRNVRLR--GKEIDVRYIQ 1248
QY 545 LSAAGTATTTDV 556
Db 1249 ---PGVASVEEV 1257

RESULT 7
US-08-617-697-9
; Sequence 9, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia

```



```

OY 87 SFAG-----KGYSLFNLKSSAEGALSVTTDKNLSTLGFSSL-TFLAAPSSTVTTPSG 140
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 300 QAAATYALPTGYTIGTIEEMNTSGAATLNTSS--GYTGTLALNTNNGAAGVYTAGAG 357
OY 141 KG-----AVKCGDLPFNNGTILFKODYCEBNGAISTKNSLKN-STGISFEGKNS 194
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 358 QNLATYTAQAANNVAVDGRANVT-----VASTGYTSGTITVGAASAAGTYSVANS 412
OY 195 ATGKKGACATGYDITNNAPTLFNNIAEAGAGINST--GNCYTGN--TSLVFS 249
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 413 TT-----TTGALAVTGGFAVT-----VAQTAGNAVTTLTQADVYVGTGNSSTAVTV 459
OY 250 ENSVTATAGN--GALSGDADVTISGNOSYTFPSGNOAVANGAIIAKLITLASGGGGIS 307
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 460 TQTAATATGATVAGRVNNAVITTDASAASATTAGKIAT-----VTLGSFQAATID 509
OY 308 FSNNTVGTAGNGAIIAAGBESLSAEGDIFPENGNAIVATTPOTTKRNSID----- 362
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 510 -SSALTYVNLSTGTSLSGI--GRGALTATPTANTLTLNNGLTGALTGAIDSEAAADGFTT 567
OY 363 ---IGSTAKITNLRAISGHSIFFDPTTANTAADSTDLNLNKADAGNSTDYSGSIVFSG 419
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 568 INAGSTASTIASLVAA-----DATLINISGDARVITTSHTAALGIVTNSV--G 618
OY 420 EKLSEDAKVADNLTSLKOPVTLTLAGNLVLRGVTLDTKGPTOTAGSSVIMDACTTLKA 479
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 619 ATLG--AELATGLV-----FTGGAGRDSILIGAT--TKAIWAGAGDITYVSATTLGA 667
OY 480 STEEYTLGLSIPVDSLEGKRVVIA--ASAASKNVALSGPILLDDNOGNAYENHDLGK 536
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 668 -----GGSV--NGDGDIVLVANVNGSFSADPAFGFETILRVAGAAAGSHN--- 713
OY 537 TQDEFVOLSAIGTAT-----TDVPAVPIVATPTPHYG----- 569
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 714 ANGFTALQLAGATAGATTTNNAVNVGLVLAAPGTGTTVLANATGSDVENLTLSSAA 773
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 570 -YQGTWGMTWVD-----DTASTPKTKTATLA-----WNTGTGLPME 605
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 774 LAAGTVALAGVETVIAATDNTTAHVDTLTLQATSAKSIYVTGAGNLNTG----- 827
OY 606 ROGPLVPNSLWGSFSDIAIOGVIERALSALTLCSDRGFMAAGVIANFLDKDKKERKRYHK 665
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 828 -----NTAVTSF--DAAAVTGT--APAVTFS-----AN-----TYGCVYTTINGG 864
OY 666 SGGVAIGAAGTQSENLSIFAFCOLFGSDKDFLAKNHTDYAGA-----FYIQHITECS 720
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 865 AGADSLTGS--TANDTIIG-----GAGADTLVYTGTDFTGTGADIDINALIGTST 916
OY 721 GFI 723
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 917 AFV 919

RESULT 9
US-09-142-648B-7
: Sequence 7, Application US/09142648B
: Patent No. 6210948
: GENERAL INFORMATION:
: APPLICANT: Smit, John
: APPLICANT: Bingle, Wade H.
: APPLICANT: No. 6210948ellini, John F.
: TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULO
: FILE REFERENCE: 08106/002002
: CURRENT APPLICATION NUMBER: US/09/142,648B
: PRIOR FILING DATE: 1999-03-30
: PRIOR APPLICATION NUMBER: PCT/CA97/00167
: PRIOR FILING DATE: 1997-03-10
: PRIOR APPLICATION NUMBER: US 07/614,377
: PRIOR FILING DATE: 1996-03-12
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 7
: LENGTH: 1026
```

```

: TYPE: PRS
: ORGANISM: Caulobacter crescentus
US-09-142-648B-7

Query Match          5.9%; Score 284; DB 4; Length 1026;
Best local similarity 23.9%; Pred. No. 5.3e-15;
Matches 187; Conservative 98; Mismatches 300; Indels 198; Gaps 40:

OY 33 PSDSFDGST--NTGYT--PKNTTGIDYTLTGDTLQNLGDSAAITKGCFSQDTBESL 86
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 243 PSSGVSGSTLSLTGCTDPLTGANNDFVAGVACAATL--TVGDT--LSGAGCTOVLMMV 299
OY 87 SFAG-----KGYSLFNLKSSAEGALSVTTDKNLSTLGFSSL-TFLAAPSSTVTTPSG 140
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 300 QAAATYALPTGYTIGTIEEMNTSGAATLNTSS--GYTGTLALNTNNGAAGVYTAGAG 357
OY 141 KG-----AVKCGDLPFNNGTILFKODYCEBNGAISTKNSLKN-STGISFEGKNS 194
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 358 QNLATYTAQAANNVAVDGRANVT-----VASTGYTSGTITVGAASAAGTYSVANS 412
OY 195 ATGKKGACATGYDITNNAPTLFNNIAEAGAGINST--GNCYTGN--TSLVFS 249
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 413 TT-----TTGALAVTGGFAVT-----VAQTAGNAVTTLTQADVYVGTGNSSTAVTV 459
OY 250 ENSVTATAGN--GALSGDADVTISGNOSYTFPSGNOAVANGAIIAKLITLASGGGGIS 307
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 460 TQTAATATGATVAGRVNNAVITTDASAASATTAGKIAT-----VTLGSFQAATID 509
OY 308 FSNNTVGTAGNGAIIAAGBESLSAEGDIFPENGNAIVATTPOTTKRNSID----- 362
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 510 -SSALTYVNLSTGTSLSGI--GRGALTATPTANTLTLNNGLTGALTGAIDSEAAADGFTT 567
OY 363 ---IGSTAKITNLRAISGHSIFFDPTTANTAADSTDLNLNKADAGNSTDYSGSIVFSG 419
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 568 INAGSTASTIASLVAA-----DATLINISGDARVITTSHTAALGIVTNSV--G 618
OY 420 EKLSEDAKVADNLTSLKOPVTLTLAGNLVLRGVTLDTKGPTOTAGSSVIMDACTTLKA 479
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 619 ATLG--AELATGLV-----FTGGAGRDSILIGAT--TKAIWAGAGDITYVSATTLGA 667
OY 480 STEEYTLGLSIPVDSLEGKRVVIA--ASAASKNVALSGPILLDDNOGNAYENHDLGK 536
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 668 -----GGSV--NGDGDIVLVANVNGSFSADPAFGFETILRVAGAAAGSHN--- 713
OY 537 TQDEFVOLSAIGTAT-----TDVPAVPIVATPTPHYG----- 569
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 714 ANGFTALQLAGATAGATTTNNAVNVGLVLAAPGTGTTVLANATGSDVENLTLSSAA 773
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 570 -YQGTWGMTWVD-----DTASTPKTKTATLA-----WNTGTGLPME 605
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 774 LAAGTVALAGVETVIAATDNTTAHVDTLTLQATSAKSIYVTGAGNLNTG----- 827
OY 606 ROGPLVPNSLWGSFSDIAIOGVIERALSALTLCSDRGFMAAGVIANFLDKDKKERKRYHK 665
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 828 -----NTAVTSF--DAAAVTGT--APAVTFS-----AN-----TYGCVYTTINGG 864
OY 666 SGGVAIGAAGTQSENLSIFAFCOLFGSDKDFLAKNHTDYAGA-----FYIQHITECS 720
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 865 AGADSLTGS--TANDTIIG-----GAGADTLVYTGTDFTGTGADIDINALIGTST 916
OY 721 GFI 723
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 917 AFV 919

RESULT 10
US-08-038-682-2
: Sequence 2, Application US/08038682
: Patent No. 5549897
: GENERAL INFORMATION:
: APPLICANT: BARENKAMP, STEPHEN J
: APPLICANT: ST. GEME III, JOSEPH W
```

```

: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
: TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/038,682
: FILING DATE: 16-MAR-1993
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: BERKSTRESSER, JERRY W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-293
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0810
: INFORMATION FOR SEO ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1536 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: US-08-038-682-2

```

```

Query Match          5.4%; Score 259; DB 1; Length 1536;
Best Local Similarity 25.5%; Pred. No. 1.3e-12;
Matches 139; Conservative 70; Mismatches 184; Indels 152; Gaps 28;

```

```

QY 26 ATAENIGPSDFDSTN-----TGTYPKN-----TTTGID--YTLTGDTTLQ--N 67
D 936 ARFKDIDNSKMLSTTTSSSTYRTILSNITNKNGDLNITNEGSDTEMOJGGDVSQKEGN 995
QY 68 L---GDSALTK-----GCFSDTTESLSFAGKGYSLF---LNK-----SS 103
D 996 LTSSDKINITKQITIKAGVGENSDSDATNNANLTIKTKELKTODLNTSGFKAETITA 1055
QY 104 AEGALSVTTDKNLSTLGFSSLTFLAAPSYYTPPSKGAVKCGDGLTFDNNGTILFKOD 163
D 1056 KDGSDLTIGNTNSADGNTAKVTFNQVSKISADGKHTLHRSKVTSGSNNT-----ED 1111
QY 164 YCEENGCAISTKMLSLKNSGSG---ISFEGNKSATGCKGACALC-TGYVDITNNAPT 218
D 1112 SSDNNAGLTIDAKVTFNNNTSHKAVSISATSGEITTKGTITNATGNEITPAQTGSI 1171
QY 219 LFSNNIAEAGGAINSTGNCITNTGNTSLV---FSENSVTATAGNGALSGDADVTISGN 274
D 1172 L-----GCISSSGSVTLITATEGALAVNSIGNTVYVTA-NSGALLTYLLASTTKGT 1221
QY 275 QSVTFSGNQ---AVANGGAIYAK-----KLTLASGGGGGGSIFSN 310
D 1222 ESVTSSQSGDIGTIGGIVEVAKATESLITQNSKTIKATGTEANVTSACTTIGG-TISG 1280
QY 311 NIYOGT-----TAGNGAISIL-----AAGECSLSAEAGDI 341
D 1281 NTFVAVTACADLTFVNGAGAEITGGAATLTITSSGKLTTEASSHTTSKAGCVNLASQGSV 1340
QY 342 TFNQNALVATTPQTRKNSIDIGSTAKITNRAISGHSIFPYDITNTAADSDTDLNLN 401
D 1341 AGSINAAVTL-----NTGTGTLTVAGSINATSGTLY-----INAKDA-----ELN 1382

```

```

QY 402 KADAG-----NSTDYSGSIVFSGEKLSEDAKVADNLTSLKQVPYTLNAGNLVKRGV-T 455
D 1383 GAALNHTYVATNANG-----SGSVIATTSRV--NITGDL---ITINGLNTISKNGINT 1433
QY 456 LDYTG 460
D 1434 VLKNG 1438

```

```

RESULT 11
US-08-302-832-2
: Sequence 2, Application US/08302832
: Patent No. 5603938
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/302,832
: FILING DATE: 16-SEP-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US pct/us93/02166
: FILING DATE: 16-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-404
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: INFORMATION FOR SEO ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1536 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
:
: US-08-302-832-2

```

```

Query Match          5.4%; Score 259; DB 1; Length 1536;
Best Local Similarity 25.5%; Pred. No. 1.3e-12;
Matches 139; Conservative 70; Mismatches 184; Indels 152; Gaps 28;

```

QY 164 YCEENG-AISTKNLSKNTSGS---ISFEGNKSATGKGAICA-TGYVDITNNAPT 218  
 Db 1112 SSDNAGLTIKAKVTNNNTSHKRAVISATSGEITTKTITTNATGNEITNAQTGSI 1171  
 QY 219 LFSNNIAEAGGAINSTGNCITGNTSLV-----PSENSVTATAGNGALSGDADVTISGN 274  
 Db 1172 L-----GGIESSGGSVLTATGALAVSNISGNTVYTA-NSGALTTLASTIGT 1221  
 QY 275 QSVTFSGNQ---AVANGALYAK-----KLTIASGGGGISFSN 310  
 Db 1222 ESVYTSQSGDIGTISGGEVEKATESLTQNSKIRATGGEANVTATGIG-TISG 1280  
 QY 311 NIYOGT-----TAGNGAISIL-----AAGECSLSAEADI 341  
 Db 1281 NTNVNTANAGDLTVNGEALNATEGAATLTSSGKLTTEASHSHITSAGQVNLASQDGSV 1340  
 QY 342 TFNGNAIVATTPQTKRNSIDIGSTAKITNLRATSGHSIFFYDPTANTADSTDTLNLN 401  
 Db 1341 AGSINAAVTL-----NTGTGLTVKGSNINATSGTLV-----INAKDA-----ELN 1382  
 QY 402 KADAG-----NSTDYSGSIVSGEKLSEDEAKVADNLTSTLKOPYTLTAGNLVLRGV-T 455  
 Db 1383 GAALGNHTVAVATNANG---SGSVIATSSRV--NITGDL--TTINGLNISKNGIN 1433  
 QY 456 LDTRG 460  
 Db 1434 VLKRG 1438

## RESULT 12

US-08-530-198-2  
 : Sequence 2, Application US/08530198  
 : Patent No. 5869065

## GENERAL INFORMATION:

APPLICANT: BARENKAMP, STEPHEN J  
 APPLICANT: ST. GEME III, JOSEPH W  
 TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
 TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Shoemaker and Matzare, Ltd  
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 STREET: Bldg. 1  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202-0286

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/530,198  
 FILING DATE: 13-DEC-1995

## CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
 NAME: BERTSTRESSER, JERRY W  
 REGISTRATION NUMBER: 22,651  
 REFERENCE/DOCKET NUMBER: JMB-1186  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 415-0813

## INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1536 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-530-198-2

Query Match 5.48; Score 259; DB 2; Length 1536;  
 Best Local Similarity 25.58; Pred. No. 1.3e-12;  
 Matches 139; Conservative 70; Mismatches 184; Indels 152; Gaps 28;

QY 26 ATAENIGSPDSFDSSTN-----TGTTPKN-----TTTGID-VTLNGDITIQ--N 67  
 Db 936 ARFDIDSKSLSTTTNSSSTYRRIISGNTNKNKNDLITHEGSPTEMOIGDVSQAKGN 995  
 QY 68 L--GDSALTR-----GCFSDTESLSPAGKYSLSF--LNIR-----SS 103  
 Db 996 LTISDKINIKQITIKAGVDGENSDPATNNANLITKTELKLTQDINISGFNKAETIA 1055  
 QY 104 AEGALSTYTTKNSLTFSSLTFLAAPSYYITPSCGAAYKCGDLTFDNNGTLLFKQD 163  
 Db 1056 KDGDLTITGNTSADGTNAKVTFFNQVDSKISADGKHVTLHSKVEYSGSNNT--ED 1111  
 QY 164 YCEENG-AISTKNLSKNTSGS---ISFEGNKSATGKGAICA-TGYVDITNNAPT 218  
 Db 1112 SSDNAGLTIKAKVTNNNTSHKRAVISATSGEITTKTITTNATGNEITNAQTGSI 1171  
 QY 219 LFSNNIAEAGGAINSTGNCITGNTSLV-----PSENSVTATAGNGALSGDADVTISGN 274  
 Db 1172 L-----GGIESSGGSVLTATGALAVSNISGNTVYTA-NSGALTTLASTIGT 1221  
 QY 275 QSVTFSGNQ---AVANGALYAK-----KLTIASGGGGISFSN 310  
 Db 1222 ESVYTSQSGDIGTISGGEVEKATESLTQNSKIRATGGEANVTATGIG-TISG 1280  
 QY 311 NIYOGT-----TAGNGAISIL-----AAGECSLSAEADI 341  
 Db 1281 NTNVNTANAGDLTVNGEALNATEGAATLTSSGKLTTEASHSHITSAGQVNLASQDGSV 1340  
 QY 342 TFNGNAIVATTPQTKRNSIDIGSTAKITNLRATSGHSIFFYDPTANTADSTDTLNLN 401  
 Db 1341 AGSINAAVTL-----NTGTGLTVKGSNINATSGTLV-----INAKDA-----ELN 1382  
 QY 402 KADAG-----NSTDYSGSIVSGEKLSEDEAKVADNLTSTLKOPYTLTAGNLVLRGV-T 455  
 Db 1383 GAALGNHTVAVATNANG---SGSVIATSSRV--NITGDL--TTINGLNISKNGIN 1433  
 QY 456 LDTRG 460  
 Db 1434 VLKRG 1438

## RESULT 13

US-08-469-880-2  
 : Sequence 2, Application US/08469880  
 : Patent No. 5876733

GENERAL INFORMATION:  
 APPLICANT: Barenkamp, Stephen J.  
 TITLE OF INVENTION: High Molecular Weight Surface Proteins  
 TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Shoemaker and Matzare, Ltd.  
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 STREET: Bldg. 1  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202-0286

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/469,880  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-516 MIS:V9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-469-880-2

Query Match 5.4%; Score 259; DB 2; Length 1536;  
Best Local Similarity 25.5%; Pred. No. 1.3e-12;  
Matches 139; Conservative 70; Mismatches 184; Indels 152; Gaps 28;

26 ATMENIGPSDFSGSTN-----TGYTPKN-----TTTGID--YTLTGDTLQ--N 67  
Db ARKEDIDNSKNLSITTTSSSTYRTIISGNTNKNGLDINTNEGSDTMQJGADVSKQEGN 995  
QY 68 L--GDSALTK-----GCFSDTESLSFAGKGYSLF--LNK-----SS 103  
Db LTISSDKINTKQTTIKAGVDGENSEDDATNNANLTKTKELKLTQDLINSGFKAEITA 1055  
QY 104 AEGALSVTTDKNLSLTFSSLFLFAPSSVITTPSGKAVKCGDLTFDNNGTILEKOD 163  
Db 1056 KDSGLTIGTNSADGTMNAKVTFNQVKDKRISADGKVTLSKVERTSGSNNT--ED 1111  
QY 164 YCEENG--AISTKLSLKNSTG--ISFEGNKSATGKGAICA--TGVDITNNAPT 218  
Db 1112 SSDNAGLITIDAKVTVNNITSHKAVSISATSGEITTKGTTINATGVEITTAQTGSI 1171  
QY 219 LFSNNIAEAGAINSTGCTTGNTSLV--FSENSVTATAGNGALSGDADVTISGN 274  
Db 1172 L-----GGIESSGSVTLTATGALAVNSISGTVTVTA--NSGALTTLAGSTIKGT 1221  
QY 275 QSYTFSGNQ--AVANGAIIYAK-----KTLFASGGGGGGSIFSN 310  
Db 1222 ESYTTSOSGDIGCTIGGIVEVKAETESLTYTOSNSKIKATVGEANVTISAGTIGG--TISG 1280  
QY 311 NIYOGT-----TAGNGAISIL-----AAGCSISAEAGDI 341  
Db 1281 NIYAVNTANAGDLIVGNAEINATEGAAITLTSSCKLTTEASSHTSKKGVNINSAQDGSV 1340  
QY 342 TFNGNALVATTPOTTKRNSIDIGSTAKITMLRAISGHSIFFEYDPTTANTAADSTDTLNLN 401  
Db 1341 AGSINANAVTL-----NTGTLTIVKGSNINMATSGLV-----INAKDA-----ELN 1382  
QY 402 KADAG-----NSTDISSSIYFSGEKLSEDAKVAADNLTSLKQPVYTLTAGNLVAKRV--T 455  
Db 1383 GAALGNTTVNANRANG-----SGSVIATITSSRV--NITGDL--ITINGNIIISKNGINT 1433  
QY 456 LDTKG 460  
Db 1434 VILKG 1438

RESULT 14  
US-08-728-470-2  
Sequence 2, Application US/08728470

Patent No. 5928651  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Matzare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,470  
FILING DATE:  
CLASSIFICATION: 424  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-MAR-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0813  
TELEFAX: (703) 415-0810  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:  
US-08-728-470-2

Query Match 5.4%; Score 259; DB 2; Length 1536;  
Best Local Similarity 25.5%; Pred. No. 1.3e-12;  
Matches 139; Conservative 70; Mismatches 184; Indels 152; Gaps 28;

26 ATMENIGPSDFSGSTN-----TGYTPKN-----TTTGID--YTLTGDTLQ--N 67  
Db ARKEDIDNSKNLSITTTSSSTYRTIISGNTNKNGLDINTNEGSDTMQJGADVSKQEGN 995  
QY 68 L--GDSALTK-----GCFSDTESLSFAGKGYSLF--LNK-----SS 103  
Db LTISSDKINTKQTTIKAGVDGENSEDDATNNANLTKTKELKLTQDLINSGFKAEITA 1055  
QY 104 AEGALSVTTDKNLSLTFSSLFLFAPSSVITTPSGKAVKCGDLTFDNNGTILEKOD 163  
Db 1056 KDSGLTIGTNSADGTMNAKVTFNQVKDKRISADGKVTLSKVERTSGSNNT--ED 1111  
QY 164 YCEENG--AISTKLSLKNSTG--ISFEGNKSATGKGAICA--TGVDITNNAPT 218  
Db 1112 SSDNAGLITIDAKVTVNNITSHKAVSISATSGEITTKGTTINATGVEITTAQTGSI 1171  
QY 219 LFSNNIAEAGAINSTGCTTGNTSLV--FSENSVTATAGNGALSGDADVTISGN 274  
Db 1172 L-----GGIESSGSVTLTATGALAVNSISGTVTVTA--NSGALTTLAGSTIKGT 1221





**THIS PAGE BLANK (USPTO)**

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 2, 2001, 03:28:38 ; Search time 76.79 Seconds  
(without alignments)  
920.562 Million cell updates/sec

Title: US-09-446-677B-4

Perfect score: 4795  
Sequence: 1 MKSQPSMLVLSSTLACFTSC.....FEVKGSSRIYVDLGGKQF 928

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

PIR 68: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4795	100.0	928	2	G86546	polymorphic outer
2	4795	100.0	928	2	G81591	polymorphic membra
3	2033	42.4	930	2	D72078	polymorphic outer
4	2028	42.3	930	2	D86546	polymorphic outer
5	2028	42.3	930	2	A81591	polymorphic membra
6	1924	40.1	928	2	E86546	polymorphic outer
7	1924	40.1	928	2	B72077	polymorphic membra
8	1855	38.7	936	2	C72078	polymorphic outer
9	1854	38.7	936	2	C86546	polymorphic outer
10	1854	38.7	936	2	B81591	polymorphic membra
11	1827.5	38.1	772	2	H86492	pmp_3 [imported] -
12	1785	37.2	928	2	H86546	polymorphic outer
13	1785	37.2	928	2	D72077	polymorphic membra
14	1785	37.2	949	2	F81591	polymorphic membra
15	1406.5	29.3	841	2	E72130	polymorphic membra
16	1403	29.3	922	2	B72131	polymorphic outer
17	1403	29.3	922	2	E86491	polymorphic outer
18	1402	29.2	922	2	F81539	polymorphic membra
19	1347.5	28.1	973	2	B86547	polymorphic outer
20	1347.5	28.1	973	2	F72076	polymorphic outer
21	1347.5	28.1	995	2	C81593	polymorphic membra
22	1273	26.5	1276	2	B86546	polymorphic membra
23	1273	26.5	1276	2	C81591	polymorphic membra
24	1218.5	25.4	1407	2	B72078	polymorphic outer
25	1109.5	23.1	1013	2	G71460	probable outer mem
26	1106.5	23.1	712	2	E86492	pmp_5 [imported] -
27	1088	22.7	445	2	E86493	polymorphic membra
28	1035	21.6	987	2	H81722	pmp_4 [imported] -
29	1009.5	21.1	359	2	C86493	

30	905.5	18.9	867	2	F81721	polymorphic membra
31	886	18.5	878	2	B71460	probable outer mem
32	816.5	17.0	424	2	D86493	polymorphic outer
33	814	17.0	497	2	A86493	polymorphic outer
34	716.5	14.9	978	2	G72076	polymorphic outer
35	711.5	14.8	978	2	B81593	polymorphic membra
36	711.5	14.8	978	2	C86547	polymorphic outer
37	698.5	14.6	1723	2	H86557	polymorphic membra
38	698.5	14.6	1723	2	E72067	polymorphic membra
39	698.5	14.6	1732	2	C81601	polymorphic membra
40	693.5	14.5	947	2	D72067	polymorphic membra
41	692.5	14.4	947	2	G86557	polymorphic membra
42	682	14.2	1672	2	C81675	polymorphic membra
43	679	14.2	946	2	C86549	polymorphic outer
44	679	14.2	946	2	D81594	polymorphic membra
45	679	14.2	946	2	C72075	polymorphic outer

## ALIGNMENTS

## RESULT 1

G86546

polymorphic outer membrane protein G family [imported] - Chlamydia pneumoniae (st

C;Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001

C;Accession: G86546

R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;

Nucleic Acids Res. 28, 2311-2314, 2000

A;Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.

A;Reference number: A86491; MID:20330349

A;Accession: G86546

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-928 <STO>

A;Cross-references: GB:BA000008; NID:g98978821; PIDN:BA96857.1; GSPDB:GN00142

A;Experimental source: strain J138

C;Genetics:

A;gene: pmp\_10

Query Match	100.0%	Score 4795;	DB 2;	Length 928;
Best Local Similarity	100.0%	Pred. No. 1.2e-256;		
Matches 928;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKSQPSMLVLSSTLACFTSCSTVFAATAENIGPDSFDGNTGTTPKMTTGIDYTLF	60	
DB	1	MKSQPSMLVLSSTLACFTSCSTVFAATAENIGPDSFDGNTGTTPKMTTGIDYTLF	60	
QY	61	GDITLQWLGDSALTKCFEDTTELSLSPACKGYSLSLNLIKSSAEGALSTYTDKNLSLT	120	
DB	61	GDITLQWLGDSALTKCFEDTTELSLSPACKGYSLSLNLIKSSAEGALSTYTDKNLSLT	120	
QY	121	GFSSIFLPLAASSVITTPPSKGAAGCGDLFPDNGNITLKRODCENGAISTKNLSLK	180	
DB	121	GFSSIFLPLAASSVITTPPSKGAAGCGDLFPDNGNITLKRODCENGAISTKNLSLK	180	
QY	181	NSTGISIFEENKSSATGKGAICATGTVDITNNTAPLFSNNIAEAGAINSTGCTI	240	
DB	181	NSTGISIFEENKSSATGKGAICATGTVDITNNTAPLFSNNIAEAGAINSTGCTI	240	
QY	241	TGNTSLVSEFNSVTATANGALSGDADVTISGNQSTFEGNOAVANGAIIYAKKLTLAS	300	
DB	241	TGNTSLVSEFNSVTATANGALSGDADVTISGNQSTFEGNOAVANGAIIYAKKLTLAS	300	
QY	301	GGGGISFSNNIYOGTTAGNGAISTLAAGECSISAPAGDITFNGNAIVATTPQTRKNS	360	
DB	301	GGGGISFSNNIYOGTTAGNGAISTLAAGECSISAPAGDITFNGNAIVATTPQTRKNS	360	
QY	361	IDIGSTAKITNLRAISGHSIFFYDPTITANTADSTDTLNLKADAGNSTDYSGSIVFSGE	420	
DB	361	IDIGSTAKITNLRAISGHSIFFYDPTITANTADSTDTLNLKADAGNSTDYSGSIVFSGE	420	

**RESULT**

**2**

81591  
polyomphic membrane protein G family CP0303 [imported] - Chlamydia pneumoniae (strain  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-May-2000 #text\_change 11-May-2000  
C:Accession: G81591  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Accession: G81591  
A:Reference number: A81500; MUID:20150255  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-928 <REA>  
A:Cross-references: GB:AEO02192; GB:AEO02161; NID:g7189226; PIDN:AAF38160.1; PID:g7189222  
A:Experimental source: strain AR39, HL cells  
/:Genetics:  
/:Gene: CP0303

Query Match	100.0%	Score 4795	DB 2	Length 926
Best Local Similarity	100.0%	Pred. No. 1,2e-256		
Matches 928	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MKSQPSWLVSTLACPFSCSTVFPAATAENIGPSPSPGSGSTNTGTGYTPKNTTGGIDYLT	60	
Db	1	MKSQPSWLVSTLACPFSCSTVFPAATARENIGPSPSPGSGSTNTGTGYTPKNTTGGIDYLT	60	
QY	61	GDITLQNLGDSAAALTRKGCFSPTTESLSFPAKGKGYSLSLFLNIKSSAEGALSVTTPDKNISLT	120	
Db	61	GDITLQNLGDSAAALTRKGCFSPTTESLSFPAKGKGYSLSLFLNIKSSAEGALSVTTPDKNISLT	120	
QY	121	GFSSLTFLPAAPSVYITTPSGKGAVKCGDGLTFDNNNGTILTFQDYTEENGGAIISTKNLSLK	180	
Db	121	GFSSLTFLPAAPSVYITTPSGKGAVKCGDGLTFDNNNGTILTFQDYTEENGGAIISTKNLSLK	180	

Qy	181	NSTGSISEBGNKSSATGKKKGALCATGTVDITNTNAPLTFENNNIAEAAGAINSGNCTI	240
Db	181	NSTGSISEBGNKSSATGKKKGALCATGTVDITNTNAPLTFENNNIAEAAGAINSGNCTI	240
Qy	241	TGNTSLVSESNKYATAGNGGALSGDADVTISGNOSVTFSGNQAIVANGALYAKKLTIAS	300
Db	241	TGNTSLVSESNKYATAGNGGALSGDADVTISGNOSVTFSGNQAIVANGALYAKKLTIAS	300
Qy	301	GGGGISFSSNNIYOGTTAGNGAISIILAAEGCSLSAEGADITFNGMAIVATTPOTTKRNS	360
Db	301	GGGGISFSSNNIYOGTTAGNGAISIILAAEGCSLSAEGADITFNGMAIVATTPOTTKRNS	360
Qy	361	IDIGSTAKITMLRAISGHSIFFEYDPTANTADSDPTLNLKKAAGNSTDVSIGSVSGE	420
Db	361	IDIGSTAKITMLRAISGHSIFFEYDPTANTADSDPTLNLKKAAGNSTDVSIGSVSGE	420
Qy	421	KLSEDEAVADNLNSTLQOPVTLTLAGNLVKRGVLDTKGFTOTAGSSVIMADAGTTLKAS	480
Db	421	KLSEDEAVADNLNSTLQOPVTLTLAGNLVKRGVLDTKGFTOTAGSSVIMADAGTTLKAS	480
Qy	481	TEEVTLIGLSIPVDSLGEGKKVVIYAASAKNNVALSGPILLLDNOGNAYEHHDLGKTODF	540
Db	481	TEEVTLIGLSIPVDSLGEGKKVVIYAASAKNNVALSGPILLLDNOGNAYEHHDLGKTODF	540
Qy	541	SFVOLSALGTATTTDVPVATPVATPTHYGYGTGGMVWDDTASPKTKATLMTNTGY	600
Db	541	SFVOLSALGTATTTDVPVATPVATPTHYGYGTGGMVWDDTASPKTKATLMTNTGY	600
Qy	601	LPNPERQGPLVNPNSLWGSFSDIOAIQGVIERSAULTLCSDRGFMAAGVANFLDKDKKGEKR	660
Db	601	LPNPERQGPLVNPNSLWGSFSDIOAIQGVIERSAULTLCSDRGFMAAGVANFLDKDKKGEKR	660
Qy	661	KYRHKSGGYAIGGAQAOTCESENLISFAFCOLFGSDKFLVAKNHTDTYAGAFYIOHITBES	720
Db	661	KYRHKSGGYAIGGAQAOTCESENLISFAFCOLFGSDKFLVAKNHTDTYAGAFYIOHITBES	720
Qy	721	GFICGLDLKLGKSMHKLVLLEGOLAYSHVNDLKTKTATPEVYGSNGNNAFNNMLCAS	780
Db	721	GFICGLDLKLGKSMHKLVLLEGOLAYSHVNDLKTKTATPEVYGSNGNNAFNNMLCAS	780
Qy	781	SHSYEYELHCFDTPAYTKLNLTYIRODSFSEKGEGRSFPDSNLFNLSLPIGVFEKFS	840
Db	781	SHSYEYELHCFDTPAYTKLNLTYIRODSFSEKGEGRSFPDSNLFNLSLPIGVFEKFS	840
Qy	841	DCDNFSYDLTISYVDDLIRNDPKCTTALVIGASWETIYANNILAROLAVRAGSHYAFSPM	900
Db	841	DCDNFSYDLTISYVDDLIRNDPKCTTALVIGASWETIYANNILAROLAVRAGSHYAFSPM	900
Qy	901	FEVYLGQFVEEYRGSSRIYNDLGGKFOF	928
Db	901	FEVYLGQFVEEYRGSSRIYNDLGGKFOF	928

```

RESULT      3
D72078
polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: D72078
C:Altman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: D72078
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-930 <ARN>
A:Cross-references: GB:AE001627; GB:AE001363; NID:94376721; PIDN:AAD18590.1; PID:g4337
A:Experimental source: strain CWL029
C:Genetics:
C:Gene: pmp-8

```

Query Match 42.4% Score 2033; DB 2; Length 930;  
Best Local Similarity 44.6% Pred. No. 1.9e-104;  
Matches 424; Conservative 158; Mismatches 326; Indels 44; Gaps 17;

QY 1 MKSQFWLVLSTLACFTSCSTVFAATNENIGPSDFGSDTNGTTPKNT--TTGIDYT 58  
Db 1 MKIPLHLKLLISSTLVTPILLSIATYGADASLSPDPSFDGAGS--TFPKSTADANGTYV 59  
QY 59 LTGDTTLONLDSALTRGCFSDTTESLSPAGKYSLSFLNKS--SAGCALSVTTDKNL 117  
Db 60 LSGNVYINDAKGKALTGCCFETTGDTLTFGKGSFSEFNVDGNSNAGAASTADAKAL 119  
QY 118 SLTGSSSLTFPLAASSVYTPPSGAVKCGDGLTFDNNGTLLFPODYCEE--NGAIST 174  
Db 120 TTFGFSNLSTFAAFGT--TVASGKSTLSSAGALMLTDGTLTFSONVSEANNGGALTT 177  
QY 175 KNLKLNKSTGISFEGNKSATGKKGALCATGVDTNNTPPLTFNNIAEAAGAINS 234  
Db 178 KTLISGNTSSITFTLSNKA--KLGAALYSSAAASISGNTGQLVFNNKKGEGGALGF 234  
QY 235 TGNCCTTGNTSLVSENSVATAGNGAL---SGDA-DVITISGQSVTFSGNOAVANG 289  
Db 235 EASSITQNSLTFSGNATDAAGKGAICYEKGTEPPLTISGKSLTFPENSSTVGG 294  
QY 290 ALVAKKLLAAGGGGGSFNSNIYOGTTAGNGAISIAGAECSLSAEAGITFNGNAIV 349  
Db 295 AICAHGLDLSA--AGPILFSNNRCGNTAAGKGAIALDSSLSISANOGDITFLGNTLT 352  
QY 350 ATPQTTRKNSIDIGSTAKITNLRAISGHSIFPYDPTANTAADSTDLNKNKADAGNST 409  
Db 353 STAPSTSRNLIYLGSSAKITNLRAAQGQSIYFDPIASNTTG--ASDVLITINQDPSNPL 411  
QY 410 DYSGSIYFSGEKLSEDEAKVADNLTSLKQPVTLTLAGNLVKRGVLTDFKFTQTAGSSV 469  
Db 412 DYSGTIYFSGEKLSEDEAKADNFTSLIKQPLALASGLTALGNVELDVNGFTQEGSTL 471  
QY 470 IMDAGTTLKASTEEVTLTGLSIPVDSLEGKRVYIAASAANKVALSPILLDNOGNAY 529  
Db 472 LMOPGTAKADTEAISTLKLVVDLSALEGKNSVSIETAGAKKTTLLSPILFODSSGNFY 531  
QY 530 ENHDLGKTODESFVQLSALGTAATT-----DPAVPTVATP--THYGOGTWGMTWVDDT 582  
Db 532 ESHINQ----AFQPLVTVFAATAADSIYIDALLTSVYQPEPHYGYGQGWEMTAD-- 585  
QY 583 ASTPKTATLAWNTGTYLPPPEROGPLVPNSLMGSPSDIOAGIVERSALTLCSDRGF 642  
Db 586 --TSTAKSGTMTVTTGYNPMPERRASVYPSLWASFDTIRTLQOIMSQANSIYQOGL 643  
QY 643 WAAVAVNLDKDKGKERKRYHKSQGYAIGGAOTCSNLSIFACOLFGSDKDFLVAKN 702  
Db 644 WASGTANFHHDKGKTNOAFRHKSGYIYVGSADFSFNISVAFCOLFGDKDLFIYEN 703  
QY 703 HTDITYAGAFYIQHTTECSGF----IGCLLDKLPGSWSHKPLVLEGOLAYSHVSDNKLTKY 758  
Db 704 TSHNYLASLYIQHRAFLGGLPMPSPGSTTMDL----KDIPILNAQLSYSTKDKMDTRY 759  
QY 759 TAYEAVGSGWNNANFNMGLASSHSY--PEYLHCFDYAPYIKMLIYIKODSESEKTEG 817  
Db 760 TSYEAGSGWNTNNGALELGGSLALYLPKEAPFOGYFPFLFOAVYSRQONFRESGEA 819  
QY 818 RSEPDNLFNLSLIGVKEFEFSDCNDSYDLTSLYVDLIRNDPKCTALAVIGASMET 877  
Db 820 RAFDGDLVNCISIVGIRLEKISDEKRNNEISLAYIGDVRKPNRSTSLMVGASWTS 879  
QY 878 YANMLARQALQVRAGSHYAFSPMEVIGQFVEYRGSSRIYVNDLGGFCF 928  
Db 880 LCKMLARQAFPLASAGSHLTLSPHVELSGEAYELRGSHIYNVDCGLRYSF 930

RESULT 4  
D86546  
polymorphic outer membrane protein G family [imported] - Chlamydophila pneumoniae (strain

C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: D86546  
R:Shirai, M.; Hirakawa, H.; Kinoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: D86546  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-930 (STO>  
A:Cross-references: GB:BA000008; NID:98978818; PIDN:BA98654.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: pmp\_8

Query Match 42.3% Score 2028; DB 2; Length 930;  
Best Local Similarity 44.5% Pred. No. 3.6e-104;  
Matches 423; Conservative 158; Mismatches 326; Indels 44; Gaps 17;

QY 1 MKSQFWLVLSTLACFTSCSTVFAATNENIGPSDFGSDTNGTTPKNT--TTGIDYT 58  
Db 1 MKIPLHLKLLISSTLVTPILLSIATYGADASLSPDPSFDGAGS--TFPKSTADANGTYV 59  
QY 59 LTGDTTLONLDSALTRGCFSDTTESLSPAGKYSLSFLNKS--SAGCALSVTTDKNL 117  
Db 60 LSGNVYINDAKGKALTGCCFETTGDTLTFGKGSFSEFNVDGNSNAGAASTADAKAL 119  
QY 118 SLTGSSSLTFPLAASSVYTPPSGAVKCGDGLTFDNNGTLLFPODYCEE--NGAIST 174  
Db 120 TTFGFSNLSTFAAFGT--TVASGKSTLSSAGALMLTDGTLTFSONVSEANNGGALTTA 177  
QY 175 KNLKLNKSTGISFEGNKSATGKKGALCATGVDTNNTPPLTFNNIAEAAGAINS 234  
Db 178 KTLISGNTSSITFTLSNKA--KLGAALYSSAAASISGNTGQLVFNNKKGEGGALGF 234  
QY 235 TGNCCTTGNTSLVSENSVATAGNGAL---SGDA-DVITISGQSVTFSGNOAVANG 289  
Db 235 EASSITQNSLTFSGNATDAAGKGAICYEKGTEPPLTISGKSLTFPENSSTVGG 294  
QY 290 ALVAKKLLAAGGGGGSFNSNIYOGTTAGNGAISIAGAECSLSAEAGITFNGNAIV 349  
Db 295 AICAHGLDLSA--AGPILFSNNRCGNTAAGKGAIALDSSLSISANOGDITFLGNTLT 352  
QY 350 ATPQTTRKNSIDIGSTAKITNLRAISGHSIFPYDPTANTAADSTDLNKNKADAGNST 409  
Db 353 STAPSTSRNLIYLGSSAKITNLRAAQGQSIYFDPIASNTTG--ASDVLITINQDPSNPL 411  
QY 410 DYSGSIYFSGEKLSEDEAKVADNLTSLKQPVTLTLAGNLVKRGVLTDFKFTQTAGSSV 469  
Db 412 DYSGTIYFSGEKLSEDEAKADNFTSLIKQPLALASGLTALGNVELDVNGFTQEGSTL 471  
QY 470 IMDAGTTLKASTEEVTLTGLSIPVDSLEGKRVYIAASAANKVALSPILLDNOGNAY 529  
Db 472 LMOPGTAKADTEAISTLKLVVDLSALEGKNSVSIETAGAKKTTLLSPILFODSSGNFY 531  
QY 530 ENHDLGKTODESFVQLSALGTAATT-----DPAVPTVATP--THYGOGTWGMTWVDDT 582  
Db 532 ESHINQ----AFQPLVTVFAATAADSIYIDALLTSVYQPEPHYGYGQGWEMTAD-- 585  
QY 583 ASTPKTATLAWNTGTYLPPPEROGPLVPNSLMGSPSDIOAGIVERSALTLCSDRGF 642  
Db 586 --TSTAKSGTMTVTTGYNPMPERRASVYPSLWASFDTIRTLQOIMSQANSIYQOGL 643  
QY 643 WAAVAVNLDKDKGKERKRYHKSQGYAIGGAOTCSNLSIFACOLFGSDKDFLVAKN 702  
Db 644 WASGTANFHHDKGKTNOAFRHKSGYIYVGSADFSFNISVAFCOLFGDKDLFIYEN 703  
QY 703 HTDITYAGAFYIQHTTECSGF----IGCLLDKLPGSWSHKPLVLEGOLAYSHVSDNKLTKY 758  
Db 704 TSHNYLASLYIQHRAFLGGLPMPSPGSTTMDL----KDIPILNAQLSYSTKDKMDTRY 759



QY 290 AIIYAKKLLIASGGGGISFSNNIVGTTAGNGAISTILAEGCSLSAEAGDITFENGNAIV 349  
 DB 290 AIIYDNLVLS--GGPILFKNNSAIDTAAPLGAIAIADSGSLSLALGGDITFEGNTVV 347  
 QY 350 --ATTPTOTKRNISIDIGST-AKITNLRASGHSIFFPDPTAANTAOSTDPTLNKMDAG 406  
 DB 348 KGASSSOTTRNSINIGNTAKIYQIARASOGNTIYFDPITTTITALSDALNNGDIA 407  
 QY 407 NSTDYSGSIVFSGEKLSEDEAKVADNLTSLKOPVTLTACNLVLRGVTLDTKGFOTAG 466  
 DB 408 GNPAYOGTIVFSGEKLSEAEAEADNLKSTIQOPLTLAGQLSLKSGVTIVAKSFSQSPG 467  
 QY 467 SSVIMDAGTTLKASTEEVTLTGLSIPVDSLGEKKVYIASAASKNALSGPILLDNQG 526  
 DB 468 STLMDAGTTLT--TADGITNNLVLNDSLSKETKATLKAQAQOTVTLSSLSLVPDPSG 526  
 QY 527 NAYENHDLGKTODESFYQLSALGTAT--TTDVPAPVTATPTHYGYGTGWTGMDVDDTAS 584  
 DB 527 NAYEDVSWNNPQVPSCLTLTADDPANHITDLADPLEKNPIHWGYGNALMSQEDTAT 586  
 QY 585 TPKTKTTLTAMTNTGYLPNPEROGPLVPSLMSGFSIDQAIQGVIERALSALTLCSDRGFMA 644  
 DB 587 --KSKATTLTWTGTGYNPNERGTLVANTLMSGFVDVRSIQQLVATKVRQSOETRGIMC 644  
 QY 645 AGVANFLDKDKKGGKRRKRRKSGGYAIGGAQTCSENLISPAFCQLGSDKDFLYAKNHT 704  
 DB 645 EGISNPFHAKSTKINKCFRHSAGYVGAATTTASDNLITPAFCQLGSDKDFLYAKNHT 704  
 QY 705 DTYAGAFYIOHITCEGSGFICGLDKLPKSGSHKPLVLEGOLAYSHVNSDLTKTKTAPEV 764  
 DB 705 SAYAASLHLQHLATLSS--PSLAKYLPFGSESEQVLEDAQSTIYSKNTMYTTOAPKG 762  
 QY 765 KGSWGNNAFNMMLGAS--SHSYPEYLHCFDTPYAPIKLNTIYIROSSEKTE--GRSFDD 822  
 DB 763 ESSWYNGCALLETALSLPHALSHGELFHAYFPFIKVEASYIHODSEKERTTLTVRSFDS 822  
 QY 823 SNLFNLSLPIGVKFEKSDCNDPSYDLTSLVDPDLIRNDPCTALVYISGASWETIYNNL 882  
 DB 823 GDLINVSVPIGITFEPERSRNERASYEATVIYADVYRKNDPCTALLINNTSMWTTGTNL 882  
 QY 883 ARQALQVRASGSHYAFSPMEYVLGOFVEVGRSSRIYVNDLGKRFQF 928  
 DB 883 SRQAGIGRAGIFAYFSPNLEVTSMLSMEIRGSSRSYNADLGKRFQF 928

RESULT 7  
 polymorphic membrane protein G family CP0306 [imported] - Chlamydomonas pneumoniae (stre  
 C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
 C:Accession: B72077; B81592  
 R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
 Nature Genet. 21, 385-389, 1999  
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A:Reference number: A72000; NCID:99206606  
 A:Accession: B72077  
 A:Molecule type: DNA  
 A:Residues: 1-928 <ARN>  
 A:Cross-references: GB:AE001628; GB:AE001363; NID:g4376730; PIDN:AAD18591.1; PID:g437673  
 A:Experimental source: strain CML029  
 R:Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heldelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Gilm, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.  
 A:Reference number: A81500; NCID:20150255  
 A:Accession: B81592  
 A:Molecule type: DNA  
 A:Residues: 1-928 <REA>  
 A:Cross-references: GB:AE002192; GB:AE002161; NID:g7189226; PIDN:AAF38163.1; PID:g718923  
 A:Experimental source: strain AR39, HL cells  
 C:Genetics:  
 A:Gene: pmp\_9; CP0306

Query Match 40.18; Score 1924; DB 2; Length 928;  
 Best Local Similarity 44.58; Pred. No. 1,9e-98;  
 Matches 421; Conservative 151; Mismatches 338; Indels 36; Gaps 18;

QY 1 MKSOFMLVLSSTACTSCS--TVEATAE--NIGSPDSFDSNTGTYTKNTTGTID-- 56  
 DB 1 MKSLHMFLLISSIALPLSLNFSAPAAVVEINLDPNFSFG--PGYTPPAQOTTMADGT 57  
 QY 57 --YTLTGDTITLONLDSALTKGCPDTESSLSPAGKYSLSFLNIKSSABCAALSVTDD 115  
 DB 58 IYNLGDVSTINMGSPALATASCKETTGNSLFGHGYQLQNDIGANGCTFNTANK 117  
 QY 116 NLSITGFSSLTFLAPSSVITTPSGKAVKCGDLPDNNCTILFKQDYCEBNGAISTK 175  
 DB 118 LLSFGFSYSLTIOTYVAT--TGTAIKSTGACSIOSNVCYFGGONFSNDNGALOGS 173  
 QY 176 NLSLKNSTGISSEFGKSSATGKGAICATGVDITNNAPLTFLENNIKSAPAGAIINST 235  
 DB 174 SISL--SLNPIITFAKNAT--QKGAISTGTGITTNNLNSASFERTANNGAIIYTE 229  
 QY 236 GNCITIGNTSLVSESVATAGNGAL--SGDAD--VTISGNQSVTESGNOAVANGC 289  
 DB 230 ASSPISNKAIISFINSVTATGATGALYCSSTAPKPVTLTSDNGELNFTGNTAITSOG 289  
 QY 290 AIIYAKKLLIASGGGGISFSNNIVGTTAGNGAISTILAEGCSLSAEAGDITFENGNAIV 349  
 DB 290 AIIYDNLVLS--GGPILFKNNSAIDTAAPLGAIAIADSGSLSLALGGDITFEGNTVV 347  
 QY 350 --ATTPTOTKRNISIDIGST-AKITNLRASGHSIFFPDPTAANTAOSTDPTLNKMDAG 406  
 DB 348 KGASSSOTTRNSINIGNTAKIYQIARASOGNTIYFDPITTTITALSDALNNGDIA 407  
 QY 407 NSTDYSGSIVFSGEKLSEDEAKVADNLTSLKOPVTLTACNLVLRGVTLDTKGFOTAG 466  
 DB 408 GNPAYOGTIVFSGEKLSEAEAEADNLKSTIQOPLTLAGQLSLKSGVTIVAKSFSQSPG 467  
 QY 467 SSVIMDAGTTLKASTEEVTLTGLSIPVDSLGEKKVYIASAASKNALSGPILLDNQG 526  
 DB 468 STLMDAGTTLT--TADGITNNLVLNDSLSKETKATLKAQAQOTVTLSSLSLVPDPSG 526  
 QY 527 NAYENHDLGKTODESFYQLSALGTAT--TTDVPAPVTATPTHYGYGTGWTGMDVDDTAS 584  
 DB 527 NAYEDVSWNNPQVPSCLTLTADDPANHITDLADPLEKNPIHWGYGNALMSQEDTAT 586  
 QY 585 TPKTKTTLTAMTNTGYLPNPEROGPLVPSLMSGFSIDQAIQGVIERALSALTLCSDRGFMA 644  
 DB 587 --KSKATTLTWTGTGYNPNERGTLVANTLMSGFVDVRSIQQLVATKVRQSOETRGIMC 644  
 QY 705 DTYAGAFYIOHITCEGSGFICGLDKLPKSGSHKPLVLEGOLAYSHVNSDLTKTKTAPEV 764  
 DB 705 SAYAASLHLQHLATLSS--PSLAKYLPFGSESEQVLEDAQSTIYSKNTMYTTOAPKG 762  
 QY 765 KGSWGNNAFNMMLGAS--SHSYPEYLHCFDTPYAPIKLNTIYIROSSEKTE--GRSFDD 822  
 DB 763 ESSWYNGCALLETALSLPHALSHGELFHAYFPFIKVEASYIHODSEKERTTLTVRSFDS 822  
 QY 823 SNLFNLSLPIGVKFEKSDCNDPSYDLTSLVDPDLIRNDPCTALVYISGASWETIYNNL 882  
 DB 823 GDLINVSVPIGITFEPERSRNERASYEATVIYADVYRKNDPCTALLINNTSMWTTGTNL 882  
 QY 883 ARQALQVRASGSHYAFSPMEYVLGOFVEVGRSSRIYVNDLGKRFQF 928  
 DB 883 SRQAGIGRAGIFAYFSPNLEVTSMLSMEIRGSSRSYNADLGKRFQF 928

RESULT 8  
 C72078

polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029)  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: C72078  
R:Kallen, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: C72078  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1936 <ARN>  
A:Cross-references: GB:AE001627; GB:AE001363; NID:94376721; PIDN:AAID8589.1; PID:9437672  
A:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: pmp\_7

Query Match 38.7%; Score 1855; DB 2; Length 936;  
Best Local Similarity 42.0%; Pred. No. 1.2e-94;  
Matches 399; Conservative 172; Mismatches 343; Indels 36; Gaps 18;

QY 1 MKSQFSLVLSSTLACTSCSTVFAATENIGPSDSFGSTNGTYPRKMT---TGIDY 57  
DB 1 MKSSVSLFFSS-IPLESSLIV-AAEVTLDDSSNNSYDGSNGT-TTFVFTTDAAGTTY 57  
QY 58 TLNGDITLQNLGD-SAAITKGFSDTESLSPAGKGYSL--SPLNKSAGALSVTTD 114  
DB 58 SLSDVSFQNAAGALGIPLAGGCFLEAGDGLTFQGNHAKLFAFINAGSSAGYASTAAD 117  
QY 115 KNLSTGESSLFLAAPSIVITPPSGKAVCGGDLTFDNNGTTLFKODYCEENGAI 174  
DB 118 KNLFFNDFSLIISCP-SLLSPTGQCALKSVGNLSLTGNSQIITFQNFSDNGVINT 176  
QY 175 KNLSTKSTGISPEGNKSSATKGGALCATGVDITNTNPTLFSNNIAEAGALNS 234  
DB 177 KNLFLSTGSPASSRNQ-AFTGKGGVYATGTTITENSPGIVSFQNLAKSGGALYS 235  
QY 235 TGNCTITGNTSLVSENSV-TATAGNGAL---SGDADVTISGNOSTVFGNOAVANGA 290  
DB 236 TDNCSITDNQVIFDGSNAEAAQAGALICCTTDTKTVLTGNKNSLFTNNATLTGGA 295  
QY 291 IYAKKTLASGGGGISFSNNIVOGTTA--GNGAISITLAGECSLSEAGDITFNGNAI 348  
DB 296 ISGLKAVISIA--GGPTLFQSNIT--SGSSAGGCGGALINIASAGELATSATSGDITFNNNOV 352  
QY 349 VATTPTTKRNSIDIGSTATITNLRAISGHISFPYDPTANTADSDTLNLKADAGNS 408  
DB 353 --TNGSTSTNAINITDAKAVTSIRATGOSIYFYPDITTPGTAASDTDLNLADANSE 410  
QY 409 TDYSGSIVSEGEKLSDEKAVADNLSTLKPVTLTAGNLVLRGYTLDTKGFOTGASS 468  
DB 411 IEYGAIVFSEGEKLSPEKAIANAVSTIRQPAVLARGDLVLRGVTVTRKDLTQSPGR 470  
QY 469 VINDAGTLKASIEEYTLGLSLIPVDSLGEGKRVVIAAASAKNVALSGPILLDNNGNA 528  
DB 471 ILMDGGTTLAKRANISLNLAVNLSSLDGTNKRAALKTEADKNISLSGTIALIDTEGSF 530  
QY 529 YEHNDLCKTODFESVOLASGATTTDVPAPVPA--PTTHYGOGTGMWTVDDAST 585  
DB 531 YEHNNLSASTYPLELTTTGANGCTITLGLSLTLLOEPETHYGOQNWOLSPNNATIS- 589  
QY 586 PKTKATLANTNGYLPNPRGQPLVPSNLSGFSIDIOAGYIERSALTLCSDRGWAA 645  
DB 590 --KISINMTTRIGYIPSPERKSNLPLNSLMGNFIDRSINOLIEFTSSSEPERELMLS 646  
QY 646 GYANPLDKKGGKRRYRKSGGYAIGGAQCSENLISAFQOLFGSKDPLVAKNHTD 705  
DB 647 GINPFYRDSMPRHRGRIHSGYALGITYATPAEDQITFAFCOLPARDNHITGKNHGD 706  
QY 706 TYAGAFYIOH---ITBESGIGLCLKPGSMH---KPLVEGOLAAYHSVNDLTKTY 758  
DB 707 TYGASLTFHHTBEGLDIANFLMKATRAPAVLSEISOIIPLSFDKFSYLTHTDHMKTY 766

QY 759 TAYEYKSGWGNNAFNNMLGASSHSYPEYLHCEPTYAPYIKLNTLYTROSFSSEKTEGR 818  
DB 767 TDNIIKSGWRNDAFCADLGSILFVIVSLYLLKEVEPFYKQVYIAHQDPERHAEGR 826  
QY 819 SFDDNSLNLSPILGKVEFEKFSDCNDSYDITLSTYVDLIRNDPKCTTALVIGASMETY 878  
DB 827 AFNKSSELINVEIPGIVFEKDSKSEKCTYDITLMTYILDAARNRKCTSLADANMAY 886  
QY 879 ANNLRALQVRAGSHYAFSPFEVLGQFVEYRSGSSRIYNVDLGGKFOF 928  
DB 887 GTNLARGFVSRANRHOVNPHEIFGQFAFEVRSSRNNTNLGSKFCF 936

RESULT 9  
C86546  
polymorphic outer membrane protein g family [imported] - Chlamydia pneumoniae (st  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: C86546  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: AB6491; MUID:20330349  
A:Accession: C86546  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-936 <STO>  
A:Cross-references: GB:BA000008; NID:98978817; PIDN:BA09653.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: pmp\_7

Query Match 38.7%; Score 1854; DB 2; Length 936;  
Best Local Similarity 42.0%; Pred. No. 1.4e-94;  
Matches 399; Conservative 172; Mismatches 343; Indels 36; Gaps 18;

QY 1 MKSQFSLVLSSTLACTSCSTVFAATENIGPSDSFGSTNGTYPRKMT---TGIDY 57  
DB 1 MKSSVSLFFSS-IPLESSLIV-AAEVTLDDSSNNSYDGSNGT-TTFVFTTDAAGTTY 57  
QY 58 TLNGDITLQNLGD-SAAITKGFSDTESLSPAGKGYSL--SPLNKSAGALSVTTD 114  
DB 58 SLSDVSFQNAAGALGIPLAGGCFLEAGDGLTFQGNHAKLFAFINAGSSAGYASTAAD 117  
QY 115 KNLSTGESSLFLAAPSIVITPPSGKAVCGGDLTFDNNGTTLFKODYCEENGAI 174  
DB 118 KNLFFNDFSLIISCP-SLLSPTGQCALKSVGNLSLTGNSQIITFQNFSDNGVINT 176  
QY 175 KNLSTKSTGISPEGNKSSATKGGALCATGVDITNTNPTLFSNNIAEAGALNS 234  
DB 177 KNLFLSTGSPASSRNQ-AFTGKGGVYATGTTITENSPGIVSFQNLAKSGGALYS 235  
QY 235 TGNCTITGNTSLVSENSV-TATAGNGAL---SGDADVTISGNOSTVFGNOAVANGA 290  
DB 236 TDNCSITDNQVIFDGSNAEAAQAGALICCTTDTKTVLTGNKNSLFTNNATLTGGA 295  
QY 291 IYAKKTLASGGGGISFSNNIVOGTTA--GNGAISITLAGECSLSEAGDITFNGNAI 348  
DB 296 ISGLKAVISIA--GGPTLFQSNIT--SGSSAGGCGGALINIASAGELATSATSGDITFNNNOV 352  
QY 349 VATTPTTKRNSIDIGSTATITNLRAISGHISFPYDPTANTADSDTLNLKADAGNS 408  
DB 353 --TNGSTSTNAINITDAKAVTSIRATGOSIYFYPDITTPGTAASDTDLNLADANSE 410  
QY 409 TDYSGSIVSEGEKLSDEKAVADNLSTLKPVTLTAGNLVLRGYTLDTKGFOTGASS 468  
DB 411 IEYGAIVFSEGEKLSPEKAIANAVSTIRQPAVLARGDLVLRGVTVTRKDLTQSPGR 470  
QY 469 VINDAGTLKASIEEYTLGLSLIPVDSLGEGKRVVIAAASAKNVALSGPILLDNNGNA 528  
DB 471 ILMDGGTTLAKRANISLNLAVNLSSLDGTNKRAALKTEADKNISLSGTIALIDTEGSF 530



QY 529 YENHDICKTOPSFVQALSTATTTPVAPYVA---TPPHYGQGTWGTWDDTAST 585  
 Db 531 YENHMLASASTYPLELETTAGANGITTLGALSTLTLOEPETHYGOQNMWLSMANATSS- 589  
 QY 586 PKRTATLANTGTYLPPEROGLVPNSLWGSFSDIOAIOGIVERSALTLCSDRGFWAA 645  
 Db 590 ---KIGSINMTRTGYIPSPERKSNPLNSLMGNFTIDISINOLITKSSGPFERELMLS 646  
 QY 646 GVANFLDKKGRKKRYRHKSGGYAIGGAOTCSENLISFAFCOLFGSDKDLVAKNHTD 705  
 Db 647 GIANFEYRDSMPTRHGFHISGYALGITTATTPAEDQLTFACQLFADRNRHITGKNHGD 706  
 QY 706 TYAGAFYIQH---ITECGFICGLDLKLPKSMH---KPLVLEGQALAYSHVNDLKTKY 758  
 Db 707 TYGASLFFHHEGLFDIANFLMGKATRAFWLSEISQIITLSPFAKFSYLTLDNMKITY 766  
 QY 759 TAYEVKSGMGNMNFMMGLGASSHSYPEYLHCFPTYAPYIKLNTLYIRQDSFSEKTEGR 818  
 Db 767 TDNSIINGSMWMDAFCADLGASLFPVISVPLLKEVEFPVVOYIYAHQDPFYERYAEGR 826  
 QY 819 SFDDSNLFNLSLPIGVNFEKFSDCNDESYDILTLSTVPDLIRNDPKCTTALVISGASWET 878  
 Db 827 AFNKSSELINVEIPIGVTFEERDSKSEKGYDILTLMYILDAYRRNPKCQSLIASANMAY 886  
 QY 879 ANNLRQALQVRAGSHAFSPMEVLQGFVEVGSRRIVYVDLGKRFQF 928  
 Db 887 GTNLARQGFVRAANHFQVNPHEIFGQFAFEVSSSRNNTNLGSKCF 936

## RESULT 10

B81591  
 polyomphic membrane protein G family CP0308 [imported] - Chlamydia pneumoniae (str  
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
 C:Accession: B81591  
 R:Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39.  
 A:Reference number: A81500; M01D:20150255  
 A:Accession: B81591  
 A:Molecule type: DNA  
 A:Status: preliminary  
 A:Residues: 1-936 <RFA>  
 A:Cross-references: GB:AE002193; GB:AE002161; MID:g7189234; PIDN:AAF38165.1; PID:g718923  
 A:Experimental source: strain AR39, HL cells  
 C:Genetics:  
 A:Gene: CP0308

Query Match 38.7%; Score 1854; DB 2; Length 936;  
 Best Local Similarity 42.0%; Pred. No. 1,4e-94;  
 Matches 399; Conservative 172; Mismatches 343; Indels 36; Gaps 18;  
 QY 1 MKSQFSLVLSSTLACTSCSTVFAATENGPDSDDGNTGTTPKNT---TGIDY 57  
 Db 1 MKSSVSWLFSS-IPRFSLSIV-AAEYTLDDSSNNSYDNGST-TFTVSTTDAAGCTTY 57  
 QY 58 TLTGDTITONIGD-SALIFKCFSDTRESISFAKGYSL--SFLNKSASGALASTVTD 114  
 Db 58 SLSDVDFQNGALGIFPLASGCFEAGGDLTFQGNQALAKAFINAGSSACTVASTAD 117  
 QY 115 KNLSTLGFSSITFLAAPSIVITPSPGKAVCGGDLTFDNGNTILFKODYCEENGAI 174  
 Db 118 KNLFFNFPSRLITISCP-SLLSPFGCALKSVGNLSITGNSQIITFQNFSSDNGVINT 176  
 QY 175 KNLISKNTSGISFEKKSATGKGAICATCTVDITNNAPLFLFSNNTAEAGAINS 234  
 Db 177 KNLILSGISQFASPSRNO-AFTGKQGGVYATGTTIENSIGIVFSQNLAKSGGALYS 235  
 QY 235 TGNCTITGNTSLVSENSV-TATAGNGAL---SGDADVTISGNSQVTFSGNQAVANGA 290

Db 236 TDNCSITDNPQVIIDGNSAMEAQAQGAICCTTTDKTIVTLGKNKLSFTNNATLTGYGA 295  
 QY 291 IYAKKLTLASGGGGISFSNNIVOGTTA--GNGAISLLAGECSLSAEGADITFNGAI 348  
 Db 296 ISGLKVSISA--GGPTLTPQSMN-SGSSAGCGGGAIAINIASAGELIASATSDITFNNNOV 352  
 QY 349 VATPQRTKRSIDIGSAKTLNLRASGHSIFFPDPTANTADSPCTLNKRADAANS 408  
 Db 353 --TWGISTRNAINIDIAKVTISIRAAATGQSIYFDPITNPGTAASDTPLNMLNADNSE 410  
 QY 409 TDYSGISYFSGEKLSEDAKAVADNLSTLKPVTTLTACNLVLRKGVTLDTGFTOTASS 468  
 Db 411 IEYGAIFYSEKSLSPTRKALIAANVTSTIRQPAYARBDLYLRGVYTFEFDLQSPSR 470  
 QY 469 VIMDAGTTLAKASTEETVLTGSLIPVDSLGEKKVYIAAASKVVALSGPILLDNOGA 528  
 Db 471 ILMGQTTLSAKEANLISGLVALNLSLDGYNKALKATEADKNISLGTALIDTEGSEF 530  
 QY 529 YENHDICKTOPSFVQALSTATTTPVAPYVA---TPPHYGQGTWGTWDDTAST 585  
 Db 531 YENHMLASASTYPLELETTAGANGITTLGALSTLTLOEPETHYGOQNMWLSMANATSS- 589  
 QY 586 PKRTATLANTGTYLPPEROGLVPNSLWGSFSDIOAIOGIVERSALTLCSDRGFWAA 645  
 Db 590 ---KIGSINMTRTGYIPSPERKSNPLNSLMGNFTIDISINOLITKSSGPFERELMLS 646  
 QY 646 GVANFLDKKGRKKRYRHKSGGYAIGGAOTCSENLISFAFCOLFGSDKDLVAKNHTD 705  
 Db 647 GIANFEYRDSMPTRHGFHISGYALGITTATTPAEDQLTFACQLFADRNRHITGKNHGD 706  
 QY 706 TYAGAFYIQH---ITECGFICGLDLKLPKSMH---KPLVLEGQALAYSHVNDLKTKY 758  
 Db 707 TYGASLFFHHEGLFDIANFLMGKATRAFWLSEISQIITLSPFAKFSYLTLDNMKITY 766  
 QY 759 TAYEVKSGMGNMNFMMGLGASSHSYPEYLHCFPTYAPYIKLNTLYIRQDSFSEKTEGR 818  
 Db 767 TDNSIINGSMWMDAFCADLGASLFPVISVPLLKEVEFPVVOYIYAHQDPFYERYAEGR 826  
 QY 819 SFDDSNLFNLSLPIGVNFEKFSDCNDESYDILTLSTVPDLIRNDPKCTTALVISGASWET 878  
 Db 827 AFNKSSELINVEIPIGVTFEERDSKSEKGYDILTLMYILDAYRRNPKCQSLIASANMAY 886  
 QY 879 ANNLRQALQVRAGSHAFSPMEVLQGFVEVGSRRIVYVDLGKRFQF 928  
 Db 887 GTNLARQGFVRAANHFQVNPHEIFGQFAFEVSSSRNNTNLGSKCF 936

## RESULT 11

H86492  
 pmp\_3 [imported] - Chlamydia pneumoniae (strain J138)  
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
 C:Accession: H86492  
 R:Shirai, M.; Hiraoka, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A:Reference number: A86491; M01D:20330349  
 A:Accession: H86492  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-772 <STO>  
 A:Cross-references: GB:BA000008; MID:g8978389; PIDN:BA98226.1; GSPDB:GN00142  
 A:Experimental source: strain J138  
 C:Genetics:  
 A:Gene: pmp\_3-2

Query Match 38.1%; Score 1827.5; DB 2; Length 772;  
 Best Local Similarity 48.2%; Pred. No. 3.1e-93;  
 Matches 379; Conservative 117; Mismatches 260; Indels 31; Gaps 11;  
 QY 158 ILFKODYCEENGAISTRKLSLKNTSGISFEKKSATGKGAICATCTVDITNNAP 217



polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029)  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: D72077  
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: D72077  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-928 <ARN>  
A:Cross-references: GB:AE001628; GB:AE001363; NID:g4376730; PIDN:AA018593.1; PID:g437673  
C:Genetics: strain CWL029  
A:Gene: pmp\_11

Query Match 37.2%; Score 1785; DB 2; Length 928;  
Best Local Similarity 40.4%; Pred. No. 8.8e-91;  
Matches 385; Conservative 173; Mismatches 344; Indels 52; Gaps 18;

QY 1 MKSOFMLVLSSTLACTSCSYFAATAENIGPSDFSGSTNTGTYPPKMTTGIDYTLT 60  
DB 1 MKTSIPWLVSVLA--FSCHLQSLANEELLSPDPSFNGINDSGTFPKTSAT--TYSLT 56  
QY 61 GDITLQNLGSAALTKCFSDTETESLSPACKGYSLSLINKSSAE--GAALSVTTDKML 119  
DB 57 GDVFEYEPGKTPILSDSCFQOTDNLTFELNGHSLTFGFIAGTAAAGAASTANKMLTF 116  
QY 120 TGFSSILFLAAPSIVTTPSGKAVKCGDLTFDNNGTILFKODYCEENGAISTRMLSL 179  
DB 117 SGFSLSFDSPTTYVT--GQGLSSAGGVNLENIRKLYVAGNFSTADGAIKGAFL 174  
QY 180 KNTGSIISFEGNKSATGKGAICATGYDITNTAPLFSNNIAAAGAINSTGNC 239  
DB 175 TGTSGDALFFNNSSST--KGAIATTAAGARIANNITGYVFLSNIASTSGAIDEGT 231  
QY 240 ITGNTSLVSENSVTAAAGGAL-----SGDADVTISGNSQVTFSGNOAVANGAIYAK 294  
DB 232 LSNKFLYFEGNAKKT--GGAICNTKASGPELITSNKTLTFAENVAVETSGAIAHAK 288  
QY 295 KLTLASGGGGSIFSPNNIVOGTTAGNGAISIILAGECSLAEAGDITTFNGNAIVAT--TP 353  
DB 289 KIALSS--GGFTEFLRNNVSSATP--KGALSIDASGELSLAEAGNTTFVRNLTITG 345  
QY 354 QTTKRNSIDIGSTAKITNLRAISGHSIFFPDPTANTAADSTDTLNLNKADAGNSTD 413  
DB 346 DTPKRNAINIGSNCKFTLEBAKNNHTIFFYDPITSE--GTSSDVLKINNAGALNPYOG 403  
QY 414 SIYFSGEKLSEDEKAVADNLTSLKOPVTLTLAGNLVLRKGYTLDTKFTPTAGSSV 473  
DB 404 TILFSGETLTADELKAVADNLTSLKOPVSLSGGKLLQKGVTLLESTFSOEAGSL 463  
QY 474 GTTLKASTEEVTLTGSLIPVDSLGEKKVVIYAASAKNVALSPIILLNOGAVENHD 533  
DB 464 GTTLSTAGSITITNLGINVDSLGKOPVSLTAKGASNKVIYSKWLMLDIEGNIYISH 523  
QY 534 LGKTODEFVOLA--LGTATTTDPA--VPVATPHYGYGTGWTWVDDTASTEPKTK 589  
DB 524 FSHQLEFLSLKITVDADVDTNVDISSILIPVAPEDPNSEYFGOGMNVNMTDTAT--NFK 581  
QY 590 TATLANTNTGLNPEROGFLVNSLWGSFSDIOAGVIERSLTICSDRGPAAGVAN 649  
DB 582 EATATWTKTFVSPBPKSALVCTLMGVFTDIRSLQDLVIGATGMEHROGFEVWSSMTN 641  
QY 650 FLUDKGEKRYKRNKSGAIGAAOCTSENLISAFCOLFGSDKPIFLAVKNNHTDYAG 709  
DB 642 FLNHTGDENNKGFPHSTGVIYIGSAHTPKDDLTFEAFCHLFAADKCFIAHNNSRYGG 701  
QY 710 AFYIQTITEC-----SGTIGLCLKLPQSWSHKPLVLBQALAYSHVSNDLKTKYT 759  
DB 702 TLFFKHSHTLPQNYLRIGRAKFSESAIEKFP--REIPLADVOVSFSDNDNMETHYT 758

QY 760 AYPEKSGWGNNAFNMALGAS-----SHSYPEYLHCFDYAPYIKLNIYIRODSFSEK 814  
DB 759 SLPESEGSWSNECIAIGIGLDLPVLSNPHP-----LKTFLPQMKVEMVYSQNSFE 814  
QY 815 TEGRSPDNLMLNLSLPIGVKFEKESDCNDPSYDLTLSTVPDLIRNDPKCTTALVISGAS 874  
DB 815 SDGSGFSTIGRLNLNLSIVYAKFVGIDGSDYSTYDLSGFEFSDYVRNRPQSPATLWSPDS 874  
QY 875 WETVYANLQALQVRAGSHYAESPMFEVLGOFEYEVYRGSSRIYNDLGRKFOR 928  
DB 875 WKIRGNLSRQAFLLRGSNNVYVNSNCELFQHYAMELRGSSRNYNVDVGTFLRF 928

RESULT 14  
F81591  
polymorphic membrane protein g family CP0302 [imported] - Chlamydia pneumoniae (s  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: F81591  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255  
A:Accession: F81591  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-949 <REA>  
A:Cross-references: GB:AE002192; GB:AE002161; NID:g7189226; PIDN:AAF38159.1; PID:g718  
C:Genetics: strain AR39, HL cells  
A:Gene: CP0302

Query Match 37.2%; Score 1785; DB 2; Length 949;  
Best Local Similarity 40.4%; Pred. No. 9e-91;  
Matches 385; Conservative 173; Mismatches 344; Indels 52; Gaps 18;

QY 1 MKSOFMLVLSSTLACTSCSYFAATAENIGPSDFSGSTNTGTYPPKMTTGIDYTLT 60  
DB 22 MKTSIPWLVSVLA--FSCHLQSLANEELLSPDPSFNGINDSGTFPKTSAT--TYSLT 77  
QY 61 GDITLQNLGSAALTKCFSDTETESLSPACKGYSLSLINKSSAE--GAALSVTTDKML 119  
DB 78 GDVFEYEPGKTPILSDSCFQOTDNLTFELNGHSLTFGFIAGTAAAGAASTANKMLTF 137  
QY 120 TGFSSILFLAAPSIVTTPSGKAVKCGDLTFDNNGTILFKODYCEENGAISTRMLSL 179  
DB 138 SGFSLSFDSPTTYVT--GQGLSSAGGVNLENIRKLYVAGNFSTADGAIKGAFL 195  
QY 180 KNTGSIISFEGNKSATGKGAICATGYDITNTAPLFSNNIAAAGAINSTGNC 239  
DB 196 TGTSGDALFFNNSSST--KGAIATTAAGARIANNITGYVFLSNIASTSGAIDEGT 252  
QY 240 ITGNTSLVSENSVTAAAGGAL-----SGDADVTISGNSQVTFSGNOAVANGAIYAK 294  
DB 253 LSNKFLYFEGNAKKT--GGAICNTKASGPELITSNKTLTFAENVAVETSGAIAHAK 309  
QY 295 KLTLASGGGGSIFSPNNIVOGTTAGNGAISIILAGECSLAEAGDITTFNGNAIVAT--TP 353  
DB 310 KIALSS--GGFTEFLRNNVSSATP--KGALSIDASGELSLAEAGNTTFVRNLTITG 366  
QY 354 QTTKRNSIDIGSTAKITNLRAISGHSIFFPDPTANTAADSTDTLNLNKADAGNSTD 413  
DB 367 DTPKRNAINIGSNCKFTLEBAKNNHTIFFYDPITSE--GTSSDVLKINNAGALNPYOG 424  
QY 414 SIYFSGEKLSEDEKAVADNLTSLKOPVTLTLAGNLVLRKGYTLDTKFTPTAGSSV 473  
DB 425 TILFSGETLTADELKAVADNLTSLKOPVSLSGGKLLQKGVTLLESTFSOEAGSL 484  
QY 474 GTTLKASTEEVTLTGSLIPVDSLGEKKVVIYAASAKNVALSPIILLNOGAVENHD 533

Db 485 GTTLLSTTAGSITTTNIGLINDSLGKOPVSLTAKGASNKVIYSGKINLIDIEGNIYESHM 544  
 QY 534 LGKTDSEFOLSA-LGTATTTDPA---VPYATPHYGYGTGWTGTVDDPASTPRTK 589  
 Db 545 FSHDQFLSLKITYADVDTINVDISSILIPYAPEDPNESEYFOGOMNNWTTDTAT--NTR 602  
 QY 590 TATLANTNGYGLPNERQGPLVPNSLWGSFSDIOALOGVIERSAITLCSDRGFMAACVAN 649  
 Db 603 EADATWTKTFEVPSPERKSAIYCNLTIMGVTDIRSLQOLLEIGATGHEHKGQGVASMTN 662  
 QY 650 FLDKDKGKERKRYRHKRGGAIGGAQOTSENLISFAKQFSGDKOFLVAKNHTDYAG 709  
 Db 663 FLHKTGDENNKGFHHTSGVYIGSATPPKODLFTFAFCHLFARDKCCFTAHNNSTRYGG 722  
 QY 710 AFYIOHTEC-----SGFIGLLDLKPLGSWSHKPLVEGOLAYSHVSNDLKTKYT 759  
 Db 723 TLFFKSHHTLOPQNYLNLGAKFSSESAIEKFP--REIPLADVOVSFHSNDNMETHYT 779  
 QY 760 AYPEVKGSGWGNNAFNMMLGAS---SHSTPEYLHCFTDYAPYIKLMLTYIRODSESEK 814  
 Db 780 SLPESEGSMSNECIAGIGLDLPFVLSNPHP----LEKTEIPQMKYEMVYVSONSFESS 835  
 QY 815 TEKSPDNLFLNLSLPIGVKFEKFSDCNDFSYDLTISYPPDLIRNDPKCTTALVIGAS 874  
 Db 836 SDRGFSTIGRLNLSTIPVGAKEVQGDIGDYTDLSGFPSVDYRNPNQSTATLWSPDS 895  
 QY 875 WETVYANILARQALQVRAGSHYAFSPMEVVLGQFVEYRGSSRIYNDLGGKFOF 928  
 Db 896 WKIRGGLSLQALFLIRGSNNYVNSNCELFQHAMELRGSSRNINVDVGIRKLR 949

RESULT 15  
 E72130  
 polyomphic membrane protein G family CP0761 [imported] - Chlamydomonas pneumoniae (stra  
 C:Species: Chlamydomonas pneumoniae; Chlamydia pneumoniae  
 C:Date: 23-Apr-1999 #sequence\_revision 23 Apr-1999 #text\_change 11-May-2000  
 C:Accession: E72130; G61541  
 R:Palman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
 Nature Genet. 21, 385-389, 1999  
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A:Reference number: A72000; MUID:99206606  
 A:Accession: E72130  
 A:Molecule type: DNA  
 A:Residues: 1-841 <RNA>  
 A:Cross-references: GB:AE001586; GB:AE001363; NID:g4376263; PIDN:AA018172.1; PID:g437627  
 A:Experimental source: strain CWI029  
 R:Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Grimm, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.  
 A:Reference number: A81500; MUID:20150255  
 A:Accession: G81541  
 A:Molecule type: DNA  
 A:Residues: 1-841 <REA>  
 A:Cross-references: GB:AE002235; GB:AE002161; NID:g7189672; PIDN:AA038561.1; PID:g718967  
 A:Experimental source: strain AR39, HL cells  
 C:Genetics:  
 A:Gene: pmp\_2; CP0761

Query Match 29.3%; Score 1406.5; DB 2; Length 841;  
 Best Local Similarity 34.3%; Pred. No. 5.5e-70;  
 Matches 330; Conservative 162; Mismatches 314; Indels 157; Gaps 23;

QY 1 MKSQFMVIVSLTACSTYFAATAENIGSDSPDSSTNGTYTPK--NTTGTGIDYT 58  
 Db 1 MKPLPLFL-L-SLTPPLSMNLGAAATTELSASNSPFGTSTTSSSKTSATDGTNYV 59  
 QY 59 LTGDTLQNL---GDSAAITKGCFSPTTES--LSFAGKGYSLSLNI-KSSAGCAAL-SV 111  
 Db 60 FKDSVIEVNPVKFGETQSTS--CFKNDAAAGDLNPLFGGGSFTFSNIDATTAAGAAIGSE 117  
 QY 112 TTDKNLSLIGFSSLTFLAPSSVITTPSGKAVKCGGDLTFDNNGTILFKQDYCENGCA 171

Db 118 AANKTYLSCGFSAISFLKSPASTYT-----NGL----- 145  
 QY 172 ISTKNLKKNSTGSIISPEKSSATGKGAICATGVIDITNNTAPTLFENNIAEAAAGA 231  
 Db 146 -----GA 147  
 QY 232 INSTGNCTITGNTSLVFSNSVTATAGNGALSGDADVTISGNOSTVSGNOVANGAI 291  
 Db 148 INVGKNSLINDNKLVLQDNFST---GDGGAINCAGSLKIANNKSLISFIQNSSSTRGAI 204  
 QY 292 YAKKTLASGGGGGSISSNNIVG---TTAGNGAISTIAEGCSLSAAGDITENGNA 347  
 Db 205 HTKNLTLSSG-----ETLFQGNAPYAGKGAIALDSQTLSTISGSGDITFEGNT 257  
 QY 348 IYATTPOTTRNSIDISTAKITNLRAISGHSIFEDYDIPANTAASTDTLNNKADAGN 407  
 Db 258 IGAT--GTVSHSAIDIGTSKAITALRAAQHTIYFDPIVTIGSTVAAIDLNTNSPDG 315  
 QY 408 STDYSGSIVSGERKLSDEAKVADNLTSLKOPVTLTLAGNLVLRGVTLDTKGFQTAGS 467  
 Db 316 NKEYTGTVSGEKLPFEAKDEKNRTSKLQNVAFKNGTVVLKGDVYLSANGFSQDANS 375  
 QY 468 SVYMDAGTYLKASTERYTLGLSIPVDSLCEGKKVYIAASAASNAVALSPILLDNGN 527  
 Db 376 KLIMDLTSLVANTESIELNLEINIDSLRNGKIKLSATAQDKRIDRPVYLAIDES 435  
 QY 528 AYENHDGKTQDF--SFYQLSA-LGTATTTDPAVPYATPHYGYGTGWTGTVDDPAST 585  
 Db 436 FYQNGFLNEDHSDGILELDAGDIYISADSRSIDAVQSP--YGYQKWTINNSTD----- 489  
 QY 586 PKTKTATLANTNGYGLPNERQGPLVPNSLWGSFSDIOALOGVIERSAITLCSDRGFMA 645  
 Db 490 --DKKATVSAKQSPNPTAQBAPLVNMLGSEFIDVRSPQNFIELTCEGAPYKRRVVA 547  
 QY 646 GYANFLDKKKGKRRYRHKRGGAIGGAQOTSENLISFAFQPLGSDKDELVAKNHTD 705  
 Db 548 GISNVLRSGRENQRFRRHVSAGVAGASTRMGDGTLISGFQALFARDKDYEMNTNFAK 607  
 QY 706 TYAGAFYIOHTECSGFIGGLDK-----LPGSMHKLPLVEGOLAYSHVSNDLKTKY 758  
 Db 608 TYAGSLRLQDASLIVSVSTLLEGGGIRELLPYVSKTILPCSTYGGQLSTGHTHRRKTES 667  
 QY 759 TAYPEV-----KSGWGNNAFNMMLG-----ASHSTPEYLHCFTDYAPYIKLMLTYIR 806  
 Db 668 LPPEPPTLSTDHTSMGWYVAGELGFRVAVENTSGRGF-----FQETPFFVQAVYAR 721  
 QY 807 QDSFSEKGTGSRFSDSNLFLNLSLPIGVKFEK-FSDCNDFSYDLTISYPPDLIRNDPKC 865  
 Db 722 QDSFVEIGALSIRFSDSHLNLNLAIPGLIKLEKRA---EQYVHVAVMYSDDVCRSNPKCT 778  
 QY 866 TALVIGASWETVYANILARQALQVRAGSHYAFSPMEVVLGQFVEYRGSSRIYNDLGGK 925  
 Db 779 TTILSNQSGSKTKGSLNARAGIYQASGFRSLCAAAELRQNFEMRGSSRSRTNVDAGSK 838  
 QY 926 FQF 928  
 Db 839 IKF 841

Search completed: October 2, 2001, 03:28:49  
 Job time: 1306 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 2, 2001, 03:31:52 ; Search time 48.09 Seconds

(without alignments)  
661.033 Million cell updates/sec

Title: US-09-446-677B-4

Perfect score: 1 MKSQFSLVLSSTLACTSC.....FEVGRSSRIYVNDLGKRFQF 928

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SWISSPROT\_39.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	580	12.1	964	1	OMPE_CHLTR
2	544.5	11.4	1034	1	OMPE_CHLTR
3	291	6.1	1249	1	190K_RICRI
4	284	5.9	1025	1	SLAP_CANCR
5	267	5.6	1250	1	YFAL_ECOLI
6	267	5.6	1656	1	OMPB_RICJA
7	248.5	5.2	1643	1	OMPB_RICPR
8	245.5	5.1	1300	1	120K_RICRI
9	245.5	5.1	1654	1	OMPB_RICRI
10	242	5.0	1645	1	OMPB_RICRY
11	238.5	5.0	2003	1	YDBA_ECOLI
12	220	4.6	1041	1	EGT2_YEAST
13	219.5	4.6	995	1	YIO9_YEAST
14	216	4.5	1039	1	AG43_ECOLI
15	215	4.5	3591	1	FHAB_BORPE
16	213	4.4	1286	1	ALDA_ECOLI
17	212.5	4.4	1306	1	MSB2_YEAST
18	212	4.4	881	1	YJH8_YEAST
19	203	4.2	1140	1	YMG6_YEAST
20	202	4.2	587	1	FIBP_ADEI2
21	202	4.2	933	1	SLAP_CAMEF
22	201.5	4.2	1150	1	APMT_PIG
23	200.5	4.2	1567	1	ICEN_XANCT
24	198	4.1	2329	1	YSH9_CAEEL
25	197	4.1	1196	1	ICEN_PSEBX
26	197	4.1	1200	1	ICEN_PSEBX
27	194	4.0	1848	1	CBPA_CLOCL
28	193.5	4.0	749	1	TROP_HUMAN
29	193.5	4.0	1325	1	YDEK_ECOLI
30	193	4.0	1322	1	ICEN_PANAN
31	192	4.0	1569	1	YFJA_ECOLI
32	191	4.0	1608	1	HLVA_SERMA
33	190.5	4.0	863	1	YEJO_ECOLI

34	190	4.0	1148	1	ICEN_PSEBX	030611 pseudomonas
35	188	3.9	1296	1	VACJ_HELPY	048247 helicobacte
36	186.5	3.9	1210	1	ICEN_PSEBX	009815 pseudomonas
37	186	3.9	1419	1	ALAL_CANAL	013368 candida alb
38	185	3.9	1258	1	ICEN_ERWHE	016239 erwina her
39	183	3.8	959	1	N100_YEAST	002629 saccharomyc
40	183	3.8	1902	1	P2P_LACIA	015293 lactococcus
41	182	3.8	598	1	YAU_F_SCHPO	010168 schizosacch
42	182	3.8	1014	1	NANH_CLOSE	029767 clostridium
43	181.5	3.8	976	1	FIBP_ADEB3	003553 bovine aden
44	180	3.8	796	1	YSH9_CAEEL	009625 caenorhabdi
45	179.5	3.7	1290	1	VACA_HELPY	055981 helicobacte

## ALIGNMENTS

RESULT	ID	OMPE_CHLTR	STANDARD	PRT	964 AA.
AC	084877	OMPE_CHLTR	STANDARD	PRT	964 AA.
DT	30-MAY-2000	(Rel. 39, Created)			
DT	30-MAY-2000	(Rel. 39, Last sequence update)			
DT	30-MAY-2000	(Rel. 39, Last annotation update)			
DE	PUTATIVE OUTER MEMBRANE PROTEIN E PRECURSOR.				
GN	PMPE OR CT869.				
OS	Chlamydia trachomatis.				
OC	Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.				
OX	NCBI_TaxID=813;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRATIN=D/WM-3/CX;				
RX	MEDLINE=9900809; PubMed=9784136;				
RA	Stephens R.S., Kaiman S., Lammell C.J., Fan J., Karathe R., Aravind L.,				
RA	Mitchell W.P., Ollinger L., Tatusov R.L., Zhao Q., Koonin E.V.,				
RA	Davis R.W.;				
RT	*Genome sequence of an obligate intracellular pathogen of humans:				
RT	Chlamydia trachomatis.*;				
RL	Science 282:754-759(1998).				
CC	- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)				
CC	(POTENTIAL).				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: AEC01360; AAC68467.1; .				
KW	Outer membrane; Signal.				
FT	SIGNAL 1 18				
FT	CHAIN 19 964				
FT	SEQUENCE 964 AA; 104703 MW; 1B998A7D2571CE2 CRC64;				
CC	POTENTIAL.				
CC	PUTATIVE OUTER MEMBRANE PROTEIN E.				

Query Match 12.1% ; Score 580; DB 1; Length 964;  
Best Local Similarity 25.4% ; Pred. No. 1.1e-24;  
Matches 262; Conservative 145; Mismatches 378; Indels 246; Gaps 49;

QY	1	MKSQFSLVLSSTLACTSCSYVPAATNENIGPSDSGDSNGTGYPPKNTTGIDITLT 60	1	ICEN_PSEBX	030611 pseudomonas
DB	1	MKRAFFFLVLSSTLACTSCSYVPAATNENIGPSDSGDSNGTGYPPKNTTGIDITLT 60	1	VACJ_HELPY	048247 helicobacte
QY	61	GDTTLQWGPSAALFKGCFDPTRESLSFACKGYSLSPLNKSSAEGAALSVTTDKNLSLT 120	1	ICEN_PSEBX	009815 pseudomonas
DB	42	NKISL-TGTHNLT-NCYIDNLR-----YIALIQ-KTPMEGAAYIT----- 81	1	ALAL_CANAL	013368 candida alb
QY	121	GFSSLTFLAAPSIV-----TPSGKGAIVCG-----GDLTDDN---GTIL 159	1	ICEN_ERWHE	016239 erwina her
DB	82	-DYLSFDQKQEIYPAKNLTPESGALIGVAPSPVTEIRDTIGVIFENNTCCRLFT 139	1	N100_YEAST	002629 saccharomyc

OY	160	KRODCEE----	NGAIIISKNLSLKNCSGISFEENKS-----	133
Db	140	WRNPAAADKIEGGAIIHQNL	YIHNHEDVYFMRNPFYVGGAISTANTYVSENOCSFL	199
OY	194	-----SATKRGGAICATG	VDITNNIAPLTFSSNNIAEAGGAINSTGNCITGNT-	244
Db	200	FMDNICIOTNTAGKGA	IYACTGSFSESNNDLEFIINN-ACCAGAIYFSP-IGSLTGNRG	257
OY	245	SLVSESEN-----	SVTPATNGGALSDDADVTISGNOS-VTEFGNOAVANGAIYAKKL	296
Db	258	NIVFNNNCFKNVETAS	SEADGGAIKYTRTLDVYTGNGRFRFESDNTTKNTGAIYAPV	317
OY	297	TLASGGGGGSI	FSNNIYOGTTAGNGAISTILAECGLSAEAGDITENGNAI-----	348
Db	318	TLVD--NGPTFI	NNIINAI-----NKGAIYIDGTSNKSISADRAHAIIFENNI	371
OY	349	VATPQQTNRNSID	IGSFAKTTNIRATISGHSIFPYDDITANTFADSDDTLINTKNADGNS	408
Db	372	TTSTNNPPRRNAI	YVASSGELLIGAGSSOMLITPYDEVMASV--VSFNK-----E	423
OY	409	TDYSGSIYFSEK	LESDRAKVAAD-----NLASTLKQPYTLFAGNLVLRKGYLIDKGTQOT	464
Db	424	ADQTSVYFSGA	IATYNS-----ADFHQRRLDKTPAPRLTNSGFLCIDHQLVYNNRTQT	478
OY	465	A-----GSSV	IMDA---GTTLAKASTEYTLGLSIPYDS-LGEBKKVYI-----	504
Db	479	GGVSLGAGVAL	SCYCKNKTGDSASNAISTIKHICLINLSIKSGAEPIPLMVEPTNNSNN	538
OY	505	-----AASA	AKNVALSPILLDNOGA-VEHNDLKTQDFSPVOLSAIGTATTTPVA	558
Db	539	YTADTAATF	SLSDVKLS---LIDDYGSPIESTDL--THAALSQPMLSISEASDNOLOS	592
OY	559	-----VPTVA	TPHYGYOGTWGMWVDYDTASTPKT-----KTATIAW	595
Db	593	ENIDESGLNV	-----HYGQGLMTWMAKTOPEPASATITPDQKANFRHTLLTLW	646
OY	596	TNTGTLPMPE	QKGPLVYPSMISGSSDIOALOGVIERSAITLCSRGW---AAGVANFLD	655
Db	647	LPAGVPSPKH	RSPILIAITLMGNM--LLATES-LKNSLEITPSPGHPWGITGGGLGMNVY	703
OY	653	KDKKEKKRYK	HRHKSQGYAIG-FAOTCSENLISFAOLFSS-----DKDFLVAKNHTDT	706
Db	704	QDPRNHFG	FHMRRSSGYAGMIAQOT---HFFSLKFSQTYTKLMERAKANNVSSKNSCQ	760
OY	707	YAGATYIDH	TECSEFGICLLDKLPGRSMKRPVLVEGOLATSHVSNDLKTYTAPREYKG	766
Db	761	GEMLFSLDE	-----GELLTKLVG-----LYSYDHNCHNHETT-QGNNLT	798
OY	767	SWGNAFPM	MMGLGASHSYPEYLHCFDT---YAPYIKLNTLYINODFSFSEKTEGSRFSD	822
Db	799	SGGTFRSOT	MGGAFFEDLP--MKRFGSGHILITAFPLDLAGIYSSLSHFTVEGATPRBSFT	856
OY	823	SN-LFNLS	LPGVKFEKFSDC--NDFSITDLSLYVPDLIRNDPKCTTALVYISGASWETYA	879
Db	857	KTPLLNVLP	IPGVK-GSEPMATHRPQAVVELAQPVLVYRQEPICIAQLLASKGIMFGSG	915
OY	880	NNIAROA	QVLR	890
Db	916	SPSSRH	AMSYK	926
RESULT 2				
OMPF_CHLTR				
AC	P38008;	STANDARD;	PRT;	1034 AA.
DT	01-OCT-1994 (Rel. 30, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	PUTATIVE OUTER MEMBRANE PROTEIN F PRECURSOR.			
OS	PMF OR CT870.			
OS	Chlamydia trachomatis.			
OS	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI TaxID=813;			

[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-L2/TW-3/CX;  
RC MEDLINE=99000809; Pubmed=9784136;  
RX Stephens R.S., Kaiman S., Lammell C.T., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RV Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans:  
RT Chlamydia trachomatis";  
RL Science 282:754-759(1998).  
RN [2]  
RP SEQUENCE OF 26-35.  
RC STRAIN-L2/43/B0;  
RV Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,  
RA Comanducci M., Christensen G., Birkelund S., Viretuo E., Ratti G.,  
RV Pallini V.;  
RL Submitted (SEP-1994) to the SWISS-PROT data bank.  
CC -I- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AE001360; AAC68468.1; -.  
KW Outer membrane; Signal.  
FT SIGNAL 1 25  
FT CHAIN 26 1034  
FT CONFLICT 34 35  
FT SEQUENCE 1034 AA; 112392 MW; 445FF4C35D463AE7 CRC64;  
  
PUTATIVE OUTER MEMBRANE PROTEIN F.  
RR -- FH (TN REF. 2)

```

Query Match Similarity      11.4% ; Score 544.5 ; DB 1 ; Length 1034 ;
Best Local Similarity      23.7% ; Pred. No. 1.le-22;
Matches 247 ; Conservative 148 ; Mismatches 408 ; Indels 239 ; Gaps 45 ;

QY      62 DITLONIGDSAAALTKGC-FSDPTESLSFAGKGYSLSFNLIKSSAEGALSVTTDKNLSLT 120
       : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      57 NIVLSNIQSNG--TGACTISGNTQTQIERSNS-----VNTTADSGCADMYTT-----SFT 104

QY     121 GFSSITFLAASSVYTTPPSKGAVYCGGDLPEDNNGTLLFQG-----DYCCEN 168
       : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     105 ASDMANLEFC-NNYCTHKKGGALRGSPPIRLANQDVLFRNNISACKKYGTGDHNENK 163

QY     169 -----GGAIST-KNLSLNSTGISFEGRKS----- 193
       : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     164 RGALYATTTTLTCNRFLAFINMSGDCGGAISADQIISIDTVGLIFENNHTLNHPY 223

QY     194 ---SATKKGGALICATGV-DITTNPAPLFESNNILAEAGAINSNGNTIGNTS-LVES 249
       : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     224 TQAENMARGAICSRDLCISISNNGSPIVEFYVN-QGGKGAISA-RVIDNNNERILFS 281

QY     250 EN-----SVTPAAGGALSGDADVITISGNO-SVTFSGNQAVANGAIYAKKLFLASGG 303
       : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     282 NMSLSHGQSSASBSNGAIQTTGGTLENNNGSIYFDSDNTATHAGALINCQIIDRD--N 339

QY     304 GGISFSNNIVOGTTAGANGAISILAAGECS--LSAECADITFGNGNAIVATTPOTT-KENS 360
       : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     340 GPVYFLNN-----SAAMGAAPNLSKPRSATNVYIHGTGDIYFNANNVVPTLDGNLGKRKL 394

QY     361 IDIGS---TAKTINLRALSHGSIPFYD-----PITANPAASDTDLNLNKMDAGSSTDYS 412
       : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     395 PHINNNEITPYTSLIGAKKDRIRFIYFDLEQWEREKENSNBPSESTSNNITPVANDETES 454

QY     413 GSIVFSGERLESDEKAVDNLTSTLKQ-PVLLTLAGNVLKRGVTLDTGK--FTOTAGSSV 469
       : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     455 GAUVFVSQMOSDRIKLMGKEHKNYIKKAPFTLLKRGTLALIEDDAELEIYNIFETQPSTLL 514

QY     470 IMDAGTTL---KASTEVTTLTGLSLPV-----DSLGGKKVLIAS 507
       : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

Db 515 ALGSGATLVKQHGKLNITNGLVILPIILKRGKSPCIRVNPQDMQNTGNGTTPSSSS 574
QY 508 AASNAVALSGPILLDD--NCGNAIENHDL--GKTODESFVQLSALGTTTDD-----V 556
Db 575 ISPMIIFNGSLIVDENYSYDSMDLSRGKAQL-----ILSIETINDGOLDNSMOS 628
QY 557 PAVYVATPTHYQGTGWTWTV-----TATLMTNTGYLPNPEROGPLVNSLMSGFS 620
Db 629 SLNLSLSPHYGIGQIMTENNITTTTILNNNSAPTSAISAEOKKTSEFTTSPSNTT 688
QY 582 TASTPKTK-----TATLMTNTGYLPNPEROGPLVNSLMSGFS 620
Db 689 TASPINIKASGSGSASNGSEVTLTKHLLVVMAPVGIYDPIRGDILANSIVHSGR 748
QY 621 DIIAQIVERSALITLSDGAFMA--AGVANFLDKDKKGEKKRYRKSGGYAIGAQT 677
Db 749 N-----MTMGRLSLIPDMSWFLQGAATTLFTKQOKRLSYHOSASASKGYTVSSQASG 801
QY 678 CSENLISFARCOLFGSDKDLVAKNHTDYAGAFYIQHTECSGFIGCLDKLPMSHSHK 737
Db 802 AHGKFLLSFQ--SSDK--MKKETNNRLSSRYTSLA--C--FEHPMFDK----- 845
QY 738 PLVLEQGLAYSHVSNLDTKYTAYPEYKSGMGNNAFMMLGASSHSYEXYLHC-FDTYAP 796
Db 846 -IALIGAACNGYTHNNRSPFYGTRKSSKGKPHST-----LGAS-----LRCLRDMSMP 893
QY 797 YIKINLTYIKQDSFS--EKTEGHSFDDSNLFLNIS-----LPYKFEKESDC-NDFS 846
Db 894 LRSIMLPFAQALFSRTEPASIRESGDLARLFTLEQAHFAVSPICIGKAGYSSTPTPLS 953
QY 847 YDLTSLVPLLRNDPCKTALVYSASWETVANNLAKOLQVAGSHYAFSPHFEVLGQ 906
Db 954 WEMELAIQPLTKRPLMLTLTIONGSMWVTYTPPLAKHSFYGR-GSHSLKFSHLKLFAN 1012
QY 907 FVEFVGRSSRIYVNDLGKRFQF 928
Db 1013 YQAEVATSTVSHYINAGCALVF 1034

RESULT 3
190K_RICRI STANDARD; PRT; 2249 AA.
ID 190K_RICRI
AC P15921;
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OC NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=90354033; PubMed=2117568;
RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
RT "A protective protein antigen of Rickettsia rickettsii has tandemly
RT repeated, near-identical sequences."
RL Infect. Immun. 58:2760-2769(1990).
CC - FUNCTION: ELICITS PROTECTIVE IMMUNITY.
CC - PFM: GLYCOSYLATED (POSSIBLE).
CC - DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF
CC AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M31227; AAA26380.1; -

```

```

DR PIR: A41477; A41477.
KM Antigen; Repeat; Signal; S-layer; Glycoprotein.
FT SIGNAL 1
FT CHAIN 29
FT DOMAIN 212
FT REPEAT 212
FT REPEAT 287
FT REPEAT 359
FT REPEAT 431
FT REPEAT 506
FT REPEAT 578
FT REPEAT 653
FT REPEAT 725
FT REPEAT 800
FT REPEAT 875
FT REPEAT 950
FT REPEAT 1022
FT REPEAT 1094
FT REPEAT 1166
SQ SEQUENCE 2249 AA; 224333 MW; ADP646C089DE087 CRC64;

Query Match 6.1%; Score 291; DB 1; Length 2249;
Best Local Similarity 24.0%; Pred. No. 2.5e-08;
Matches 149; Conservative 78; Mismatches 245; Indels 148; Gaps 28;

QY 9 VLSTTACFTSCSTVFAIAEN---IGPSDFSGSTVGTTPKNTTGTIDYLTGDI 65
Db 234 VIRATTKTLTNAASVLTLTNANAVLTGAINNTGQDVGVL---NLNGALSGVTGDI-- 287
QY 66 QNLGDSALNRKGCSDPTSELSFAGKYSLSFLNIRKSSABEALSYTDKNLSLTGSS 125
Db 288 -----GNTNSLATISVAGATVTLG-----GAVKATYTK--LTDASA 323
QY 126 TFLAPSSVITTPSGKAVGCGDLTFDNNGTILFKODYEENGALSTKLSKNSTGS 185
Db 324 VKFTNP--VVVT-----GALDNTGNA--NNGIVTF-----TGNSTVGTGVTNMLAT 367
QY 186 ISFEGKSSATGKKGAICATGYVDITNNTAPILFSN-----NIAEAGAINSTG 236
Db 368 V---NVGAGILQYQGGVAVNANFTINTDNASAVFTNPVVVTGALDNTGNANNGIVFTG 423
QY 237 NCTTT---GNTSLVFSNSVTATAGNGALSGDADVTISGNQVTFSGNQAVANGA 293
Db 424 NSTVTDIGINTNALATVNVGATTLGGAIVKATYTKLTNNAASVLT---TLTNAHAVLT 478
QY 294 KULTIASSGGG-GISFNNITVQGTAGNGAISILAECSLSAEADITFNGNAIVATT 352
Db 479 GADNTTGGDNVGLNLNGALSQVT-GNIGNTNSLA---TISVAGATLGAIVKATY 533
QY 353 PQTTRNSIDIGTAKITNLRAISGHSIFRYDPTTANTADSTDTLNLN-----K 402
Db 534 TKLT-----DAASAVKFTN-----PVVVTGALDNTGNANNGIVTFTGNSVT 575
QY 403 ADGNSTDY-----SGSIVFSGEKLSEDAKVADNLTSTLKQVTVTFAGNLVLRGVT 456
Db 576 GDIGNTSLATISVGAATLGLGAVIKATYTKLT--NAAVLT---TLTNAHAVLTGAI 630
QY 457 DTGK-----FTQTAG-----SSVINDACT-TLKASTEEVTLTGLSIPVSL 496
Db 631 TTGGDNVGLNLNGALSQVTDIGNTNSLATISVGAATLGLGAVIKATYTKLTNNA 690
QY 497 GEGKRVYIASSAKSNVALSGPILLDNOGNAYENHDLGKTODESFVQLSALGT----- 550
Db 691 KFTNPVVVTALIDSTGNANG--IVFTGNSVTGDTGINTNALATVNVGA-GTATLGA 746
QY 551 ---ATTVDPAVPVATPT 567
Db 747 VIRATTKTLTNAASVLTLTN 766

RESULT 4
SLAP_CAVCR

```



ID SLAP\_CAUCR STANDARD; PRT: 1025 AA.  
AC P35828; Q46015;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-OCT-1994 (Rel. 30, last sequence update)  
DT 30-MAY-2000 (Rel. 39, last annotation update)  
DE S-LAYER PROTEIN (PARACRYSTALLINE SURFACE LAYER PROTEIN).  
GN R5AA.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_Taxid=76;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.  
RC STRAIN-AFCC 19089 / CB15;  
RX MEDLINE=93007489; PubMed=1393820;  
RA Gilchrist A., Fisher J.A., Smit J.K.;  
RT "Nucleotide sequence analysis of the gene encoding the Caulobacter  
RT crescentus paracrystalline surface layer protein.";  
RL Can. J. Microbiol. 38:193-202(1992).  
RN [2]  
RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.  
RC STRAIN-AFCC 19089 / CB15;  
RX MEDLINE=89008089; PubMed=3049545;  
RA Fisher J.A., Smit J.K., Agabian N.;  
RT "Transcriptional analysis of the major surface array gene of  
RT Caulobacter crescentus.";  
RL J. Bacteriol. 170:4706-4713(1988).  
RN [3]  
RP CHARACTERIZATION.  
RC STRAIN-CB15;  
RX MEDLINE=98292737; PubMed=9620954;  
RA Awram P., Smit J.K.;  
RT "The Caulobacter crescentus paracrystalline S-layer protein is  
RT secreted by an ABC transporter (type I) secretion apparatus.";  
RL J. Bacteriol. 180:3062-3069(1998).  
CC - FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY  
CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.  
CC - SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A  
CC S-LAYER WITH HEXAGONAL SYMMETRY.  
CC - MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE  
CC SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A  
CC SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF062345; AAC38665.1; -  
DR HSSP: P01549; 2MCM.  
DR InterPro: IPR001343; -  
DR Pfam: PF00353; hemolysincabind; 1.  
DR PRINTS: PR00313; CABNDNGRPT.  
KW Cell wall; S-layer; Calcium-binding.  
FT INIT MET 0  
FT 0  
SO SEQUENCE 1025 AA; 98209 MW; Afc8B519820B1A5F CRC64;

Query Match 5.9%; Score 284; DB 1; Length 1025;  
Best Local Similarity 23.9%; Pred. No. 2.2e-08;  
Matches 187; Conservative 98; Mismatches 300; Indels 198; Gaps 40;

OY 33 PDSFSGST---NTGYT---PKNTTGTIDYTLTGDTLQNLGDSALITGCGSDTTESTL 86  
DB 242 PSSGVSGSTLSTLTGTDLTGNTANNDTFVAGEVAGATL-TVGD-TLSGAGTDLVLMV 298  
OY 87 SFG-----KGVSLFELINKSSAEGALSVTEKNSLTGFSSTL-FLAPSSVITTPSG 140  
DB 299 QAAAVTALPTGTISGLETMWNTSGAAILTNISS--GVGTALNTMTSGAQTVTAGAG 356

OY 141 KG-----AYKGGDLTFDNNGTILFKODYCEENGALSTKNSLCK-STGSIPEGNKSS 194  
DB 357 QNTATTAQAQANNVAVDGRANT-----VASGVSGTITTVGANSASGTVSVVANS 411  
OY 195 ATGKKGALCATGTVDITNTTAPLFSNNIAEAGAGINST---GNCTIGN--TSLVFS 249  
DB 412 TT-----TTGAIATVGTAVT-----VAQAGNANVTTLTQADVTYGNSTTAAVY 458  
OY 250 ENSVTATAGN--GGAISGADVITISGNSVTFSGNOAVANGAIVAKKILIASGGGGIS 307  
DB 459 TQTAATAGATVAGRVNGANTITDSAAASATITAGKIA-----VILSFGAATID 508  
OY 308 FSNNVQGTAGNGAISIILAGECSLSAEGDITFNGAIVATTPQTKRNSID----- 362  
DB 509 -SALTTVNLSGTGTSGIGI--GREGALTATPPANLTITLVNMLTTTGAITDSEAAADGFTT 566  
OY 363 ---IGSTAKITNLRASIGHSIFEPYDPTANTADSDTTLNLKNADAGNSPDYSGSIYFSG 419  
DB 567 INTAGSTASTIASLVAA-----DATTILISGDARVITISHAALLTGITVNSV---G 617  
OY 420 EKSEDEAKYADNLITSLKQPTLTAGNLVLRKGVTLDPKGFQTAGSSVIMDAGTTLKA 479  
DB 618 ATLG-----AELATGLV-----FTGAGRDSILLGAT--TKAIYMGAGDDVTVSSATLGA 666  
OY 480 STEEVTITGLSTIPVDSLGECKKVIA---ASAASKNVALSGLILLDNGCNATENDDLK 536  
DB 667 -----GGSV-----NGGDDTDLVANVNGSSFSADPAFGFETLVRVAGAAAGSHN--- 712  
OY 537 TODESFVQLSALGTATT-----TDVPAPVPTVATPTHYG----- 569  
DB 713 ANGFTALQGTAGATTFTVAVANNGLITVLAAPTGTITVLANATGSDVFNLTSSAA 772  
OY 570 -YQGTGMVTWD-----DTASTPKTKTATLA-----WTNTGYLPNPE 605  
DB 773 LAAGTVALAGVETVNIATPTNTTAHYDTLTQATSAKSIIVTGNAGNLNTNG----- 826  
OY 606 RQGPVLPNSLMGSEFSDIQAIQVIERSALTLCSDRGWAAGVAFDLKKKGKRRYRK 665  
DB 827 -----NIAVTSF--DASAVTGT--APAVTFV-----AN---TTVGEVWTIRGG 863  
OY 666 SGGVATGAAGTQSENLSISPAFQLGSDKDFLVAKNHTPTVGA-----FYIOHTICS 720  
DB 864 AGADSLTGS--TANDITIG-----GAGADTLVYTGIDTFTGCGADIFDINALGTST 915  
OY 721 GFI 723  
DB 916 AFV 918

RESULT 5  
ID YFAL\_ECOLI STANDARD; PRT: 1250 AA.  
AC P45508; P45507; P45506; P39441; P76468; P77487;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1997 (Rel. 35, last sequence update)  
DT 01-OCT-2000 (Rel. 40, last annotation update)  
DE HYPOTHETICAL 131.2 KDA PROTEIN IN UBIG-NRDA INTERGENIC REGION  
DE PRECURSOR.  
GN YFAL.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_Taxid=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
Science 277:1453-1474(1997).



DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
 GN OMPB.  
 OS Rickettsia japonica.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 CC NCBI\_Taxid=35790;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-YH;  
 RA Uchiyama T.;  
 RT "Sequencing of the gene encoding the protein romp B of Rickettsia  
 RT japonica";  
 RT Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
 CC SIMILARITY).  
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AB003681; BAA20138.1; .  
 CC Antigen: S-layer; Cell wall.  
 CC FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.  
 CC FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.  
 CC FT DOMAIN 528 533 POLY-SLY  
 CC FT SEQUENCE 1656 AA: 168097 MW: 3132A69C9D5999F CRC64:  
 SO

Query Match 5.6%; Score 267; DB 1; Length 1656;  
 Best local Similarity 21.0%; Pred. No. 3.5e-07;  
 Matches 194; Conservative 114; Mismatches 316; Indels 298; Gaps 44;

QY 10 LSTLACTSCSTVPAPATAENICPSDFSGTNTGYTPKNTTGTIDYTLTGDTITLONLG 69  
 DB 376 IDDTAFKTRASIVAITONSNEFTTD-FGNLAQVTPDITMLTG--NFTGD--ANNPG 429  
 QY 70 DSAALRKGCSDPTTESIFGKGYSLSLFNKKSAGGALSTYTDNLSTLGGSSILFLA 129  
 DB 430 NTAGV-----ITRA-----ANGTLASASADANVAAT--NNITALE 462  
 QY 130 APSSVITTPSGKAVKCG---DLTFDNNGTILFRQDYCEENGGAISTNLSLKSTGS 185  
 DB 463 A-----SGGVVQLDSGHTHLELRLGNAGSVFKLADGVINGKYNQIVLYGVGLAAGA 514  
 QY 186 ISEGGKSSATGK--KGALICATGYDITNNATPLF--SNNTAEAGCAINSTGCTT 241  
 DB 515 IITDGS-ARTTGIDGNGGGAGALQSTILANDAKRTLLGAGNITISANGGFINFOAN--G 570  
 QY 242 GNTSLFSENSVT-----ATAGNG-----GATSGADATVISGNOSTVPSGN----- 282  
 DB 571 GTIKLTSTONNIYVDDLAITDQTVVDASSITNAOTLTITIGTIGICANNITLGFQFNI 630  
 QY 283 ---QAVANGAIIYAKRLTLASGGGGISFSNN-----I 312  
 DB 631 GSKTTLNGVAINELVY--GNNGSVGFANHTYLLTRTTNMAAGCKIIFNPVNNNTTL 688  
 QY 313 VOGTTRAGN-----GGAISITLAAGCSLSAERAGDITFN--GNA 347  
 DB 689 AAGTNGSAAPLAEINFGSKGARADTVLVNVEGVNLYATNTITTTAANGSEFENAGGN 748  
 QY 348 IVATT---PQTKRNSI--DIGSTAKITMLRAISGHSIFEPDITANTADSDTTLN-- 401  
 DB 749 IYSGIVGGQGGKFNFTVALDNGTTFYKFLGNATFNNGT-----TIANSTLIQISGN 799

QY 402 KADAGNSTDYSGSIVFSGE-----KLSDEAKYADNLSTLKPVTLAGNLVK----- 451  
 DB 800 TADFIASADGTGIVEFYNTGPINVTLLKQAVPV-----NLKQIYTSYGGPNVYNEGNA 854  
 QY 452 ---RGVTLDRKGFQTQAGSSVIM-----DAGT-----TLKASTEERTVLGLSIP--- 492  
 DB 855 GNYHAMTDITAFENSSLAGVFLPSCGIPENDAGNTFTPLIKSTVGNETAEGRSPSVIV 914  
 QY 493 --VDS-LGEKKV-----VLAASAKKNVALSGPILLDNOGNAEYEHDKTQDSFVQ 544  
 DB 915 SGVDSVADQVIGDQNNIYGLIGLSDNGIIVNATLYAGIGTINNQ-----962  
 QY 545 LSAIGVTTT-DVPAVYVTAFTHYGQGTGWTGMDTSTFKRTATLATNTGYLPN 603  
 DB 963 ---GTYTLGGVPPNP---GTYVGLGTGIG-----ASKRQYFTFDYNNLNGI-- 1005  
 QY 604 PERQGLPVNSLWGSFSDIOAIOGVIERSALTLCSDRGFMAAGVANFLDKRKKRKYR 663  
 DB 1006 -----IATN-----TTIND-----GVT-----1017  
 QY 664 HKSGVAIGGAOTCSBNLISFAFCQ-----LFGSDKDFLV---AKNHTDTYAGA 710  
 DB 1018 VTTGGIAGGIAGTDGKITLGSVNGNANVRADGIFNSTSMIVYTKRANNGTYYLGN 1077  
 QY 711 FYIOHTREC-----SGFIGLIDKLP--GSMS-----HKPLVLEQOLAYSH 749  
 DB 1078 AIFGNIGDSPTTPVASVFSGSNNGAGLKNYISQVYIDFGRYNGIYNSNVILGSGTTAIN 1137  
 QY 750 VSNDLTKRYTAYPEVKGSGNN 771  
 DB 1138 GKIDLLTNTLTFFAGGTSTWGN 1159

RESULT 7  
 OMPB\_RICPR STANDARD; PRT; 1643 AA.  
 AC Q53020; Q9ZCM0;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, last sequence update)  
 DT 30-MAY-2000 (Rel. 39, last annotation update)  
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
 DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
 GN OMPB OR SPA OR SPA OR RP704.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 CC NCBI\_Taxid=782;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-BREINL;  
 RA MEDLINE=91045972; PubMed=2122457;  
 RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;  
 RT "Characterization of the gene encoding the protective paracrystalline-  
 RT surface-layer protein of Rickettsia prowazekii: presence of a  
 RT truncated identical homolog in Rickettsia typhi.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BREINL;  
 RA Moron C.G., Yu X.J., Walker D.H.;  
 RT "Sequence analysis of ompB of Rickettsia prowazekii.";  
 RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RA MEDLINE=9039499; PubMed=9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Sichelitz-Ponten T., Alsmark U.C.M., Podowski R.M., Neslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kuriland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of

RT mitochondria.";  
RL Nature 396:133-140(1998).  
RN [4]  
RP PARTIAL SEQUENCE.  
RC STRAIN-BREINL;  
RX MEDLINE=92114896; PubMed=1370573;  
RA Ching W.M., Carl M., Dasch G.A.;  
RT Mapping of monoclonal antibody binding sites on CNR fragments of  
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
RT prowazekii.";  
RL Mol. Immunol. 29:95-105(1992).  
RN [5]  
RP IDENTIFICATION OF CLEAVAGE SITE.  
RX MEDLINE=92104668; PubMed=1729180;  
RA Hackstadt T., Messer R., Cieplik W. Jr., Peacock M.G.;  
RT Evidence for proteolytic cleavage of the 120-kilodalton outer  
RT membrane protein of rickettsiae: identification of an avirulent mutant  
RT deficient in processing.";  
RL Infect. Immun. 60:159-165(1992).  
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
CC LAYER WITH HEXAGONAL SYMMETRY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL: M37647; AAA26390.1; ALT\_INIT.  
DR EMBL: AF161079; AAD42234.1; -.  
DR EMBL: AJ235273; CAI15140.1; -.  
KW Antigen; S-layer; Cell wall.  
FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.  
FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.  
FT VARIANT 257 257 V -> A (IN STRAIN BREINL).  
FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).  
FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).  
FT CONFLICT 178 179 AA -> VC (IN REF. 1).  
FT CONFLICT 191 201 TIOEAPLTGA -> INSRSSSHLVS (IN REF. 1).  
FT CONFLICT 212 212 T -> I (IN REF. 1).  
FT CONFLICT 313 313 Q -> L (IN REF. 1).  
FT CONFLICT 1104 1104 D -> G (IN REF. 2).  
FT CONFLICT 1123 1123 T -> S (IN REF. 2).  
SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;

Query Match 5.2%; Score 248.5; DB 1; Length 1643;  
Best Local Similarity 21.9%; Pred. No. 3.6e-06;  
Matches 228; Conservative 118; Mismatches 391; Indels 303; Gaps 52;

QY 6 SWVLSTLACTSCSTVPAATAE-----NIGPSDSFDGS-----TNTGYTPPKNT 51  
Db 601 SSTLNQOTLTINGSIGTVANTYTLAQLNIGSSKTLINAGDAVINELVIENNSVOLNHN 660  
QY 52 TTGIDYTLTGIDITLQNLGDSAAALTKGCFSDTT-----ESLSFACKGYSL-S 96  
Db 661 T-----YLITITIAANOGQIIVAADPLNTYTLADGNTNGSAENPLSTIIFAKKANADS 716  
QY 97 FLNIKSSAEGALSVTT-DKNLSLTGFSSLTFLAAPSVT-TPSGGAVKCGGDLTFDN 154  
Db 717 ILVWGVNLYANNITTDANV-----GSLHFRSGHSIYSGVGGQGHKL-NLILDN 770  
QY 155 NGTILFRODCEGNGAISTKN-LSLKN-----STGISIF-EGKSSATGKKG 201  
Db 771 GTTVKFLGDTTENGKRIEKSILQISNNYTTDVSADTQILEFVNDPITVTLNKG 830  
QY 202 AICATGVTDTNNTAPTLFS--NNIAEAGAIINSTGNCITGNTSLVFSN-----SV 253

Db 831 APEGLKQVIISGPGNIYFNGIHWGIAANSISEKNSLSTSLPSCPTLDVIFI 890  
QY 254 TATAGN-----GALSD-----ADVTISGNQSVTEG- 281  
Db 891 KSTFGNCTVDNFPAPIVVSIGDSIMINQIIGDKKNIALSLSDNSIYVANNTLSGI 950  
QY 282 -----NQAFA-----NGAIYAKKL----- 296  
Db 951 RTKKNQGTVTLSGMNPNTIYGLIENGSPRLKQVETTTDYNLGSIIANNVTINDY 1010  
QY 297 -TLASGGGGISFENNIVOGTTAGNGA-----ISILAEGCSLSAEDITFNG 345  
Db 1011 VTLTGTIAGIDFPAKTKTLLSGVNGNANVRPDSFTSPRSKIYATQ-----ANKTVIYLG 1066  
QY 346 NAIYAT-----TPQTYR-----NSIDIGSTAKITNR-----A 374  
Db 1067 NALVSNIGSLDTPVASVFTGNDGACLOGNIYSONIDFG-TYNMLTLNSNVILGGTTA 1125  
QY 375 ISGHSIFDYDPIFANTFA-ADSTDFLNKADAGNSTDYSGSIVSGEKLSEDA----- 427  
Db 1126 INGE-----IDLNNLLIFANGSTWGDNTSISTLVNSSGNI-----GOVVAIEDAQVNAAT 1178  
QY 428 -----KVADNLSTLK--QPVTLAAGNLVLRGVTLDTKGFTQTAGSSVIMDAGTTLKA 479  
Db 1179 TGTITIKIODNANANFSGTQATYTLIOGG--ARFNGITLAPRFANT-GSNIFVKY-ELIRD 1234  
QY 480 STEEYTLGLS--IPVDSLGEKKRVIAASAASKNALSGPILLDNOGNAYENHDLGKT 537  
Db 1235 SNOQYVLTFRNDVYLVNTVAVGSAIANAPGVSONISR-----CLESTNTAAYNNMLLAKD 1290  
QY 538 QDSFVQLSLGTFITTDVAVPVPVAPPTHXGYGTGTMVWDPTASTPKTKATLMTN 597  
Db 1291 PSDVATFVGAIADTSA--AVTVN-----LNDT-----OKTODLLSNR 1337  
QY 598 TG---YLPNERQGLPVPNSLMSGFSFDIQAIOGIERSALTLCSDRGFMAAGVANFLDKD 654  
Db 1328 LGTLRYLSMAETSD--VAGSATGAVSSGDEA-----VSYGWAKPFNINAEQD 1374  
QY 655 KKGERRRYRHKSGGYAIGAAQCTSENLISFAFCQLGSDKDFLVAKNHHDTVYAGAFYIQ 714  
Db 1375 KKGIGIAGYKAKTGVVY-GDPTLASDMLMIGAIGITKTD-----IKHQYKRG----- 1422  
QY 715 HITGCSGFIGCLDKLPQGSMS-HKPLVLEQOLASHSHSNDLKTYYA-----PE 763  
Db 1423 DKTDIND-----LSFSLVGSQOLVNFPAQGNALFT--LNKYSKSORFPESNGKMSKQI 1476  
QY 764 VKSGWGNNAF--NMMLGASSHSYPEYLHCFDTPYATIKLNTYIROSDFSEKTE----- 816  
Db 1477 AAGNYDMTGTGNGILFQYDYNAMNVL-----YTPMAGLSYLSKSNENKYEKTTVANR 1531  
QY 817 -GRSFDSDNLNLSLPYGVK 835  
Db 1532 INSKFSD-----RVDLIYGA 1547

RESULT 8  
120K\_RICRI ID 120K\_RICRI STANDARD; PRT; 1300 AA.  
AC P14914;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE 120 KDA SURFACE-EXPOSED PROTEIN.  
GN P120.  
OS Rickettsia rickettsii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiidae; Rickettsia.  
OX NCBI\_TaxID=783;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-R;  
RX MEDLINE=90136087; PubMed=2515418;

RA Gilmore R.D. Jr., Joste N., McDonald G.A.:  
 RT "Cloning, expression and sequence analysis of the gene encoding the  
 RL 120 kD surface-exposed protein of *Rickettsia rickettsii*.";  
 CC Mol. Microbiol. 3:1579-1586(1989).  
 CC  
 CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS  
 CC RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A  
 CC S-LAYER WITH HEXAGONAL SYMMETRY.  
 CC  
 CC -1- DISEASE: RICKETTSIA RICKETTSII IS THE CAUSATIVE AGENT OF ROCKY  
 CC MOUNTAIN SPOTTED FEVER (RMSF).  
 CC  
 CC -1- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPTPOSES  
 CC CONFERRING ANTIGENICITY TO THE PROTEIN.  
 CC  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X16353; CAA34402.1; -  
 CC  
 CC DR PIR: S07575; S07575.  
 CC  
 CC Antigen: Glycoprotein; S-layer.  
 CC  
 CC FT CARBOHYD 7 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 66 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 86 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 103 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 147 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 268 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 330 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 375 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 415 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 424 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 430 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 436 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 444 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 515 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 547 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 593 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 655 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 698 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 710 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 799 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 800 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 826 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 844 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 861 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 879 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 920 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 926 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 1116 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 1128 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 1140 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 1146 N-LINKED (POTENTIAL).  
 CC FT CARBOHYD 1211 N-LINKED (POTENTIAL).  
 CC  
 CC SQ SEQUENCE 1300 AA; 132801 MW; E09E52C3F647243D CRC64;

Query Match 5.1%; Score 245.5; DB 1; Length 1300;  
 Best Local Similarity 21.8%; Pred. No. 3.9e-06;  
 Matches 198; Conservative 111; Mismatches 308; Indels 293; Gaps 45;

Db 110 -SGAGVQLSGTHAAELRLNAGSIFELADGVYNGKVNQTAALVGALAGTTLDG--- 165  
 Qy 194 SAT-----GKKGCAICATGVDTNNTPALF--SNNAIAGAIASTNGCTTGTSL 246  
 Db 166 SATITGIGNAGGA-KALQKITLANDAKKTLLTGANIIAGGGTIDLOAN---GGIKL 221  
 Qy 247 VFSNSVT-----ATANGC---GALSGDADVTISG-----NOS 276  
 Db 222 TSTQNNIVVDLAIATDQGVADVASSLTNAQTLITIKGIGTIGANNTKLGQFNIGSSKT 281  
 Qy 277 VTSNGNA-----YANGAL-----YAKTLTASGGGGISISNNIVQTT--AGN-- 320  
 Db 282 VLSNGVAINELVIGNDGAQVAFADHTYLLTRTNAAGQKIIIPNVNNGTTLAAGTNLG 341  
 Qy 321 -----GCAISIIAGECSLSAEGADTFN--GNAIVAT-- 352  
 Db 342 SATNPALAEINFGSKGVAVDVLAVNGEVENIXAINITTDANVCSFVNAGTNIVSGTG 401  
 Qy 353 -POTTRKNSIDI--GSTAKITNLRAISGHSIEFFDPTANFAADSTDITLNL-KADAGNS 408  
 Db 402 GQGGKNFNTVALENGTTVKFLGNATFNGNT-----TIANSTLQIGNTADCVAS 452  
 Qy 409 TDYSGSIVSGE-----KLESDAKVADNLTSLKQPVTLTACNLVYK-----RGVT 455  
 Db 453 ADGTGVEFVNTGPTVTLNKAAPV-----MALKQTVSGPENVYINIEIGNAGNHGAV 507  
 Qy 456 LDTKGFTQTAGSSVIM-----DAGT-----TLKASTEVEVTLGLSIP-----YDS-- 495  
 Db 508 TDTIAPENSSLGAVFLPRGIPENDAGNTMPLEIKTSYVGKTKAKGPDVPSVYVLGDSVI 567  
 Qy 496 ----LGEGRKVVYIASAASKNVALSGPIL-----LLDQGNAYENHDIGKTQDPSFQV 544  
 Db 568 ADQGVIGDQNNIVIGLIGSDNGIIVATLYAGISTLNNQGT-----VT 612  
 Qy 545 LSAIGTATTDVAVNPVAPPTHYGVQGTGNTVVDSTAPTKATTLANTNGVLPN 604  
 Db 613 LSG-----GVENTP-----GTVYGLGTGIG-----ASKFQVFTTIDYNNIGNT-- 651  
 Qy 605 EROGPVPLNSLW--GSEFSDIOATQGVIERSAITLCSDRGFMAAGVANFLDKDKGEKKY 662  
 Db 652 ----IATNATINDGVYTTGGIAGIGFDGKITLGSVNG-----NGNVR-- 690  
 Qy 663 RHKSGGVAIGGAQTSSENLISFAFQLFSGDKDFLYAKNHTDTYGAFTYQHTTEC--- 719  
 Db 691 ----FADGILLNSTS-----MIGTK-----ANNGVTYVYLGNAFVGNIGSDP 730  
 Qy 720 -----SGFICLLDKLP--GSMS-----HKPIVLEGOLAVSHVSNDKKTYTAY 761  
 Db 731 VASVRFETGSDSGAGLOGNITSQVIDFETVNLGIYVNSNIIIGGCTTAINGKIDLVNTLTF 790  
 Qy 762 PEVKGSGWNN 771  
 Db 791 ASGTSTWGN 800

RESULT 9  
 ID OMPB, RICRI STANDARD; PRT; 1654 AA.  
 AC 053047;  
 DT 30-MAY-2000 (rel. 39, Created)  
 DT 30-MAY-2000 (rel. 39, Last sequence update)  
 DT 30-MAY-2000 (rel. 39, Last annotation update)  
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)  
 DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
 GN OMPB.  
 OS *Rickettsia rickettsii*.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.  
 OX NCBI\_Taxid=783;  
 RN [1]  
 RP SEQUENCE FROM N.A.

CC STRAIN-R;  
 RX MEDLINE=92167802; PubMed=1724278;  
 RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;  
 RT "The 120 kilodalton outer membrane protein (omp B) of Rickettsia  
 rickettsii is encoded by an unusually long open reading frame:  
 RT evidence for protein processing from a large precursor.";  
 RL Mol. Microbiol. 5:2361-2370(1991).  
 RN [2]  
 RP SEQUENCE OF 279-1654 FROM N.A.  
 RC STRAIN-R;  
 RX MEDLINE=90136087; PubMed=2515418;  
 RA Gilmore R.D. Jr., Joste N., McDonald G.A.;  
 RT "Cloning, expression and sequence analysis of the gene encoding the  
 RT 120 kD surface-exposed protein of Rickettsia rickettsii.";  
 RL Mol. Microbiol. 3:1579-1586(1989).  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL: X16353; CAA34403.1; -  
 KM Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1 1333  
 FT DOMAIN 1334 1654 32 KDA BETA PEPTIDE.  
 FT 1181 1188 POLY-THR.  
 SQ SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

Query Match 5.1%; Score 245.5; DB 1; Length 1654;  
 Best Local Similarity 21.8%; Pred. No. 5.4e-06;  
 Matches 198; Conservative 111; Mismatches 308; Indels 293; Gaps 45;

QY 21 STEVAATAENIG-PSDSFDSSTNTGTYTPKNTWTGIDYTLTGDT--LQNLGDSAAATLKG 77  
 DB 379 TTAKTAAAKVTITQDSNFGNTDGNLAQIKVNA-ITLGNFTGDSANGNAGV--- 434  
 QY 78 CFSDTESLSFAGKGYLSLFLNKSABGAALSVTTDKNLSTGFSSITFLADSSVIT 137  
 DB 435 -----ITFDANGTLESASADANAVT--NNITATIA----- 463  
 QY 138 PSKGAVKCGG---DLFPDNGTILFKQDYCENGGALISFKNLSTKNSGTSISFEGNKS 193  
 DB 464 -SGAGVYQLSTHAARLRLGNAGSIFKLADSTVINGKVNQALGALAACTITLDG-- 519  
 QY 194 SAT-----GKKGAICATGYDINNTAPTLF--SNNTAEGAAGAINSTGCTTGNTSL 246  
 DB 520 SATITGIDGNAGGA-ALQRTITLANDAKKTLITLGANINIGAGGTIDLDQAN---GGTIKL 575  
 QY 247 VFSNSVST-----ATANGG---GALSGDADVTISG-----NQS 276  
 DB 576 TSTONNTIVVFDLAIADQGVVDASSITVNAQTLITNGKLTIGTIGANKKTLGQFNIGSSKT 635  
 QY 277 VTFSGNQA-----VANGGAI-----YAKKTLTASGGGGGGSFNNIVOGTT--AGN--- 320  
 DB 636 VLSNGVAINELVIGNDGAVQFAHDTLLITRTVNAAGGKLTFFVNVNNGTTLAAGTNLG 695  
 QY 321 -----GGAISILAGECSLSAAGDITFN--GNAIVATF-- 352  
 DB 696 SATNPILAEINSGSVNDVLANGBGVNLATNITTTDANVGSFVNAGSTNIVSGV 755  
 QY 353 -PQTKRNSIDI--GSTAKITNLRAISGHSITFYDPTTANTADSTDTLIN-KADAGNS 408  
 DB 756 GQDNKRTVALENGTVKFKLGNAFPNGNT-----TIANSTLQIGGNTYADCVAS 806

QY 409 TDYSGSIVSGE-----KLSEDEAKVADNLTSTLKQPTVLTAAGNLVK-----RGVT 455  
 DB 807 ADGTGIVEFVNTGPTTTLNKOAFV-----NALKQITVSGPKNVYIEIGNAGHHRAV 861  
 QY 456 IDTKGFTQTAGSSVYM-----DAGT-----TKASTEEVTLTGLSIP-----VDS-- 495  
 DB 862 TDTIAFENSSLGAVFELRGIPFNDAGNTMPLTIKSTVGNKTAKGFDVPSVVLGVDSVI 921  
 QY 496 -----LGEGRKRVVTAASAKSNVALSGPIL-----LDDNGMAYENHDLKTDQSFVQ 544  
 DB 922 ADGGVITGQNNNIVIGLGSDDNGIYVNTATLVAGISTLNNGGT-----VT 966  
 QY 545 LSAIGATTTPVPAVPYAFPHGYGTGWTWDDASTPKRTATLANTGTLPNP 604  
 DB 967 LSG-----GVPNTP-----GTVIGLGTGIC-----ASFKQVTFPTDNNLGNIT--- 1005  
 QY 605 ERQGPLVPNSLM--GSFSDIALQGVIERSAJLTLCSDRGFMAAGVAFDLDRKKEKRY 662  
 DB 1006 -----IATNATINDGVYTTGTGIAGIGFDGKITLGSVNG-----NGNVR-- 1044  
 QY 663 RHKSGVAIGCAQTCSENLISFAFCQLFGSDKDFLVAKNHTDYAGAFYIOHTEC--- 719  
 DB 1045 -----FADGILSNST-----MIGTTK-----ANNGTVYILGNAPVGNIGDSDTP 1084  
 QY 720 -----SGFIGGLDLKP--GSMS-----HKPLVLEGQLAYSHVNDLKTXYAY 761  
 DB 1085 VASVRFSGSDGAGLQGNVYSQVDFGTYNLGVNSNITLGGGTTAINGKIDLVNTLTLE 1144  
 QY 762 PEVKGSMGNN 771  
 DB 1145 ASGTSTWGN 1154

RESULT 10  
 OMPB\_RICTY  
 ID OMPB\_RICTY STANDARD; PRT; 1645 AA.  
 AC P66989;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
 DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
 GN SLP.  
 OS Rickettsia typhi.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_Taxid=785;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WILMINGTON;  
 RX MEDLINE=94040787; PubMed=8224886;  
 RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;  
 RT "Cloning and sequence analysis of the gene encoding the crystalline  
 RT surface layer protein of Rickettsia typhi.";  
 RL Gene 133:129-133(1993).  
 RN [2]  
 RP PARTIAL SEQUENCE.  
 RC STRAIN=WILMINGTON;  
 RX MEDLINE=92114896; PubMed=1370573;  
 RA Ching W.M., Carl M., Dasch G.A.;  
 RT "Mapping of monoclonal antibody binding sites on CNB- fragments of  
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
 RT prowazekii.";  
 RL Mol. Immunol. 29:95-105(1992).  
 RN [3]  
 RP IDENTIFICATION OF CLEANAGE SITE.  
 RX MEDLINE=92104668; PubMed=1729180;  
 RA Hackstadt T., Messer R., Cleplak W., Peacock M.G.;  
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer  
 RT membrane protein of rickettsiae: identification of an avirulent  
 RT mutant deficient in processing.";

```

RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L04661; ABA4987.1; -.
KW Antigen; S-layer; Transmembrane; Cell wall.
FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1 1354 1645 32 KDA BETA PEPTIDE.
FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
FT CONFLICT 657 657 H -> N (IN REF. 2).
FT CONFLICT 842 842 V -> I (IN REF. 2).
FT CONFLICT 1071 1071 G -> A (IN REF. 2).
FT CONFLICT 1306 1306 G -> S (IN REF. 2).
SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

Query Match 5.0%; Score 242; DB 1; Length 1645;
Best Local Similarity 22.7%; Pred. No. 8.3e-06;
Matches 195; Conservative 93; Mismatches 313; Indels 258; Gaps 42;

QY 6 SWLVSLLACFTSCSTVFATAE---NIGPSDSF-----DGSTNT--G 44
DB 602 SSLNNOFLINSGITIGANTKILGRFNVGSSKTIINAGDVAINELMENDGSVHLTHN 661
QY 45 TY-----TPKNTTGTI-DYITLG--DITLONLGDSAALTKCFESDPT 83
DB 662 TYLTITKINANOGRKIIVADPIINTDPLADGTMLGSAESPLSIHPTAKANG----- 715
QY 84 ESLSPACKGVSLSLFLNKSSAEGALSVTDKNLSLGFSLFTLPAPSSVIT-TPSGKG 142
DB 716 DSLIHICKGVNLVANNI-----TTTDANY--GSLHRSCTGTSVSGTGGQQ 760
QY 143 AVKCGDLTFDNNGTILFKODYCEENGCAIS-----TKNLSLNKSGISGF-E 189
DB 761 GLKL-NLLIIDNGTYVFLDITFNGCTIEKGSILDISSNYITDHEIADNCTLEFVN 819
QY 190 GKKSSATGKKGGA-----ICATGTV--DITNNTAPLFSNNIA--EAAGCA--- 231
DB 820 TDTITVTLNKGAFVGLKQVWVSGPNIAPFNEIGNGVAHAIAVDISFENASIGASLFL 879
QY 232 -----INST-GMCTITGNLSVFSNSVATATAGNG-----ALSDADAVTI 271
DB 880 LSGTPLDLVTIKSTVGNGTVDNENAPLIVSGDSMTINNQVIGDKNIIALSIGSDNSI 939
QY 272 SGNOGVTFSG-----NOAVA-----NGAIYAKKL----- 296
DB 940 TVNSNTLYAGIRTKTNQGVITLSGIPNPNPCTIYGLGENDPKLKQVFTDYNNLGS 999
QY 297 -----TLASGGGGISFSNNIYOGTTAGNG-----AISIAGAECSSAEEA 338
DB 1000 IIAITNTIINDVLTGIGTDFDCKITLGSINGNANVAFVDRTEFSHPTSMIYSTRANO 1059
QY 339 GDTFNGNAIV-----ATPQOTKR-----NSIDGSTAKTNLR----- 373
DB 1060 GIVTYLGNALVNGISDIPVASRFTGNDGVLGQGNHSQNDDE-TYNLTILNSDVI 1118
QY 374 -----AISG-----HSIFEDPITANTADSTDTLNLINKADAGNSTYSGSIYSGE 420
DB 1119 LGGGTAINAEIDLLRNLLIFANGSTWGNNTSLSTTLVNSGNGVQIYAIEGAQV-NAT 1177
QY 421 KLSDEKAVADNLSTLK--QPVTLTAGNLVLRGVTLDTKGFQTAGSSVIMDACTTLK 478

```

```

DB 1178 TTGTTTIKIDNNANANSGTQTYTLIOGG--ARENGTIGAPNEFDVT-GNNIFVYELIRD 1234
QY 479 ASTREYV-TLGLSLIPVDSIGCKKVVYAASAASKNVALLSGPILLDQGNAYEHDLGKT 537
DB 1235 ANQDYVLTFTNDVLTNTVTTAVGNSAINAPGVHONIA-----ICLESYDTAAYNNMLAK- 1289
QY 538 QDFSEVOLSLAGTATTTDVPVAPTPVATPTHYGVGTGWTGWDVDPASTPKTKATLMTN 597
DB 1290 -DSSDV-ATFIGALATDTGAAVATVN-----LNDT-----QKTQDILGNR 1327
QY 598 TG---TYLPNPROGPLYPNLSIMGSFSD-IOAIGCIERSALTLLCSDFGFAAGYANFLDK 653
DB 1328 LGAIRYLSNSTEAD--VGSGSTGAVSSGDEALIDV-----SYGVNAKPEPYNIAEQ 1375
QY 654 DKRGEKRRYRHSKSGVAIG 672
DB 1376 DKRGLAGYKAKTAGVYVG 1394

```

```

RESULT 11
ID YDBA_ECOLI STANDARD; PRT; 2003 AA.
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.
GN YDBA OR B140L/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blatter F.R., Plunkett G., Iii, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alida H., Bada T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampel K., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Mada C., Yamamoto Y., Horuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=92190338; PubMed=1665988;
RA Moszer I., Glaser P., Danchin A.;
RT "Multiple IS insertion sequences near the replication terminus in
RT Escherichia coli K-12."
RL Biochimie 73:1361-1374(1991).
CC -1- SIMILARITY: TO S. TYPHIMURUM ORF NEAR CYS6 (AC P25928).
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. THE GENE CODING FOR
CC BETWEEN AMINO ACIDS 839 AND 840.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	DR	EMBL; AE000237; AAC74483.1; ALT_SEQ.	
CC	DR	EMBL; AE000237; AAC74487.1; ALT_SEQ.	
CC	DR	EMBL; D90778; BAA15009.1; ALT_SEQ.	
CC	DR	EMBL; D90778; BAA18880.1; ALT_SEQ.	
CC	DR	EMBL; D90779; BAA18881.1; ALT_SEQ.	
CC	DR	EMBL; X62680; -: NOT_ANNOTATED_CDS.	
CC	DR	Ecogene; EG11307; ydbA.	
CC	KW	Hypothetical protein.	
CC	FT	CONFLICT 489 489 I -> V (IN REF. 2).	
CC	FT	CONFLICT 495 495 I -> V (IN REF. 2).	
CC	FT	SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;	
CC	QY	Query Match 5.0%; Score 238.5; DB 1; Length 2003;	
CC	QY	Best Local Similarity 20.3%; Pred. No. 1.7e-05;	
CC	QY	Matches 171; Conservative 125; Mismatches 340; Indels 205; Gaps 36;	
CC	QY	26 ATAENIGPSDFSGSTWGTGTYTPKNTTGTGIDYTLTGDTITLQNLGDSAAALTKGCFSDTTES 85	
CC	QY	306 STTNGGTGQTQINGDDATANNKTTVDGKDSGT-EI---NGNKGVIQDG-----D 354	
CC	QY	86 LSFAGKGYLSFLNLIKSSAEGALSVTDK---NLSTGFSSLTFLAAPSSVITTPSGKG 142	
CC	QY	355 LDVSGGGHGDITGDSATVDNKGKMTWTDPSIGIQVDGQAVNNCESAI--TNGGTG 412	
CC	QY	143 AVKGGDLTFDNG--TILFKQ-----DYCENGGAISTKMLSKNKGSTISFEQKKSAT 196	
CC	QY	413 TQINGDDATANNKTTVDGKDSGTGTEIAGNNGKVIQDGLDVGSGGHGIDITGD--SAT 470	
CC	QY	197 GKKGATCAT-----GTVDITNTTATPLSNNTAEAGGAINSTGN 237	
CC	QY	471 VDNKGTMTVDPESIGIQDQDAIVNNEGESTITNGTGTQINGNDATA-----NNSGK 525	
CC	QY	238 CTITGNTSLFSESVTATAGNGALSGDADVTISG-----NOSVTFSGNQAVAN 287	
CC	QY	526 TTVDG-----KDTGKFIAGNIGIVNLDSLTVVGAGHVENIGDNGTVNNKGDIIVSD 579	
CC	QY	288 GGAIYAKKLTASGG-----GGISFNINVOQTAGNGAISILAAEGCSLSAEGADI 341	
CC	QY	580 TGSIG-----VLINGEGATVSTGDNVSNATGFSITNSGKVSAGSMQVDFSTGVDL 635	
CC	QY	342 TFGNAIVATPQT-----TRNSIDIGSTAKITNLRAISGHSIFFFDPIANTAAD--ST 395	
CC	QY	636 NGNNSVTAAKDLKVVYQKATGINVSGDANTVN---ITGNVLVDKDK-TADNAAEYFFD 691	
CC	QY	396 DTNLNKADAGNSTDYSGSI-VFSGEKLSEDEAKVADNLSTLKQPVTLTAGNLVLKRG- 453	
CC	QY	692 PSVGIVNYGSDNNVTLGKLTIVYSDSEVTSRQSLFDGSAEKTSGLVVGIGDNTVNNGG 751	
CC	QY	454 -----VTLDTKGTGTTA-----GSSVIMDAGTTL----- 477	
CC	QY	752 LELIGKNALADGSOVTSRLATGYSTSVIVVSGESSVYLNNGDITISGEFFLPGFAGVIRVQ 811	
CC	QY	478 -KASTE---EVTLT-----GLSIPVDSL-----EGKKV-----VIAASASKN 512	
CC	QY	812 DKALLEIGSGATLTQMDDISDFEHGHTTVEIQNLGFATVGTGNTGINSGTISLQNGKD 871	
CC	QY	513 VALSGPILLDNOGNAYENHDL-GK-TQDFSEVQLSAGTATT---TDVPAVPTVATPT 566	
CC	QY	872 PAPSPIVLLATNGSATNAGTIKGTQEQHSVFNKYSTGHSNFIKNDVSSITGLVAQS 931	
CC	QY	567 H-----YGGQGTWGTWDDTASTPKTKTATLAWNTGYLPNPERQGLVFN 613	
CC	QY	932 NSTIINTDSGIIDLYG-RGSVGLAIADSTA-----ENQKKTLD 970	
CC	QY	614 SLWGSFSDIQAIQGVIERSAITLCSDFWAGVAVNFLDKDK-----GEKKRYRHK 665	
CC	QY	971 SMWVDANDTTAMRDIASNSAIDFTGTVG---VGTDSYSGAKNATINQLGGVITINAG 1027	

QY	666	SGGYAIGAQAATCTSENLSFAFCQLFGSKDFLVAKNHTDTYAGAFYIQHITECSGFIGC 725	
Db	1028	AGMAAYGASNTVINQGTINL---EKNGYDLSLAA---NVLGMAVVEHGTAINDQTGV 1080	
QY	726	L 726	
Db	1081	I 1081	
RESULT 12			
EGT2_YEAST			
ID	EGT2_YEAST	STANDARD;	PRT; 1041 AA.
AC	P42835;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	01-NOV-1997 (Rel. 35, Last annotation update)		
DE	EGT2 PROTEIN PRECURSOR (EARLY G1 TRANSCRIPT 2).		
GN	EGT2 OR YNL327W OR N0320.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=S288C / FY1676;		
RC	MEDLINE=95373280; PubMed=7645347;		
RA	Maftahi M., Nicaud J.-M., Levesque H., Gaillardin C.;		
RT	"Sequencing analysis of a 15.4 kb fragment of yeast chromosome XIV		
RT	identifies the RPD3, PAS8 and KRE1 loci, five new open reading		
RT	frames";		
RL	Yeast 11:567-572(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A. AND FUNCTION.		
RC	MEDLINE=96251274; PubMed=8668141;		
RA	Kovacech B., Nasmyth K., Schuster T.;		
RT	"EGT2 gene transcription is induced predominantly by Swi5 in early		
RT	G1";		
RL	Mol. Cell. Biol. 16:3264-3274(1996).		
CC	-I- FUNCTION: SEEMS TO BE INVOLVED IN THE CORRECT TIMING OF CELL		
CC	SEPARATION AFTER CYTOKINESIS, AS SEPARATION OF MUTANT DAUGHTER		
CC	CELLS IS DELAYED. COULD EITHER BE AN ENZYME NECESSARY FOR GLUCANS-		
CC	DEGRADATION OF THE CELL WALL AT THE NECK REGION BETWEEN MOTHER AND		
CC	DAUGHTER CELLS OR A REGULATORY PROTEIN CONTROLLING THIS METABOLIC		
CC	STEP. EXCLUSIVELY EXPRESSED BETWEEN THE END OF MITOSIS AND EARLY		
CC	G1; INACTIVATED BEFORE CELLS PASS START.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
CC	EMBL; Z46259; CAA86371.1; -		
DR	EMBL; Z71603; CAA96259.1; -		
DR	SGD; S0005271; EGT2		
KW	Glycoprotein; Repeat; Signal; Cell cycle; Cell division; Mitosis.		
FT	SIGNAL	1	20
FT	CHAIN	21	1041
FT	DOMAIN	200	203
FT	DOMAIN	381	384
FT	DOMAIN	388	395
FT	DOMAIN	490	493
FT	DOMAIN	586	589
FT	REPEAT	457	492
FT	REPEAT	577	606
FT	REPEAT	613	647
FT	REPEAT	716	745
FT	REPEAT	773	802
FT	REPEAT	811	840
FT	REPEAT	849	886
FT	REPEAT	887	924



```
Db 314 IS-----GDLTLTGKVIATGVVVAAGAKLLLDGDKYSFSA---DLKVYGDLLVKKSKET 366
QY 356 TKRNSIDI-GSTAKIT-----NLRAISGHSEIFFDPITANTAADSTDTLNLKADAGNST 409
Db 367 YGTEFDISENFDTGNFNAESAATSAISYFFTPSFDNSGDI--SLSLSKSKGGEVT 424
QY 410 --DYSGSIVF-----SKEKLSED-----EAKVADNLTSTLKQPVY----- 442
Db 425 FSPYSNGAFSFSNALILNGSVSGLQRRDDTEGSVNGNEINLDSYVIVEPVSGKGTV 484
QY 443 -LPAGNLVLKRGVTLDTKGTOTAGSVIMDAGTTLKASTEEVTLTGLSTPVSGLGKK 501
Db 485 NIISGNLYLHPDT-----FT--GQTVFKGEGVLAVDPTETNAT-----PIPVVGYTGK 532
QY 502 VVIAAASAKNVALSGPI-LLLDNQGNAVYENHDG-----KTQDES----- 541
Db 533 NQIAITADITALSYDGTGVTATQGNRQSFAGTGFSSDFSVSGIFAGAYAYVLYNY 592
QY 542 --FVOLSALGTATTTDVPVPTVATPHYGYQGTGWTGTTVDATSTPKTKTATLAWNTG 599
Db 593 NGVVATSAASSTASGASAVTGST-----SFGASVTGSTASTSFGASVTGSTASTS 644
QY 600 Y 600
Db 645 F 645

RESULT 14
AG43_ECOLI
ID AG43_ECOLI STANDARD; PRT; 1039 AA.
AC P3180; P76360; P75614; P97241; Q46771;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ANTIGEN 43 PRECURSOR (AG43) (FLUFFING PROTEIN).
GN FLUJ.

OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
[1] SEQUENCE FROM N.A.
RP STRAIN-K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh H., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA Sivasubraman S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamanoto Y., Horiuchi T.;
RT "A 460-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ML 308-225;
RA Henderson I.R., Owen P.;
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PRELIMINARY SEQUENCE OF 53-78.
RC STRAIN=ML 308-225;
RX MEDLINE=89291704; PubMed=2661530;
RA Caffrey P., Owen P.;
```

```
RT RT
RT 43, a unique protein complex associated with the outer membrane of
RL J. Bacteriol. 171:3634-3640(1989).
RN [5]
RP SEQUENCE OF 53-63.
RC STRAIN-K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
RN [6]
RP GENE NAME.
RX MEDLINE=97257509; PubMed=9103983;
RA Henderson I.R., Meehan M., Owen P.;
RT "Antigen 43, a phase-variable bipartite outer membrane protein,
RT determines colony morphology and autoaggregation in Escherichia coli
RT K-12.";
RL FEMS Microbiol. Lett. 149:115-120(1997).
CC -I- FUNCTION: CONTROLS COLONY FORM VARIATION AND AUTOAGGREGATION. MAY
CC FUNCTION AS AN ADHESIN.
CC -I- SUBUNIT: HETERODIMER OF THE PERIPHERAL MEMBRANE PROTEIN (ALPHA
CC CHAIN) ANCHORED TO THE INTEGRAL OUTER MEMBRANE PROTEIN (BETA
CC CHAIN).
CC -I- SUBCELLULAR LOCATION: OUTER MEMBRANE ASSOCIATED.
CC -I- SIMILARITY: TO ADHESIN AIDI-1 AND TO BORDETTELLA PERTACTIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
```

```
EMBL; AE000291; AAC75061.1; ALT_INIT.
DR EMBL; D90838; BAA15825.1; ALT_INIT.
DR EMBL; D90839; BAA15832.1; ALT_INIT.
DR EMBL; U24429; AAB47869.1;
DR Ecogene; EG12686; flu.
KW Outer membrane; Signal.
FT SIGNAL 1 52
FT CHAIN 53 551 ANTIGEN 43 ALPHA CHAIN.
FT CHAIN 552 1039 ANTIGEN 43 BETA CHAIN.
FT VARIANT 2 2 K -> N (IN STRAIN ML 308-225).
FT VARIANT 41 42 SL -> FF (IN STRAIN ML 308-225).
FT VARIANT 46 46 T -> K (IN STRAIN ML 308-225).
FT VARIANT 157 157 W -> L (IN STRAIN ML 308-225).
FT VARIANT 188 188 V -> F (IN STRAIN ML 308-225).
FT VARIANT 303 305 ATN -> STI (IN STRAIN ML 308-225).
FT VARIANT 320 320 A -> T (IN STRAIN ML 308-225).
FT VARIANT 372 372 E -> Q (IN STRAIN ML 308-225).
FT VARIANT 493 493 N -> V (IN STRAIN ML 308-225).
FT VARIANT 497 497 H -> Y (IN STRAIN ML 308-225).
FT VARIANT 585 585 S -> N (IN STRAIN ML 308-225).
FT VARIANT 709 709 E -> K (IN STRAIN ML 308-225).
FT VARIANT 721 721 M -> T (IN STRAIN ML 308-225).
FT VARIANT 751 753 GHL -> SHF (IN STRAIN ML 308-225).
FT VARIANT 803 803 S -> P (IN STRAIN ML 308-225).
FT VARIANT 815 815 A -> V (IN STRAIN ML 308-225).
FT VARIANT 824 824 C -> S (IN STRAIN ML 308-225).
FT VARIANT 829 835 LNLVHTS -> MNLINYA (IN STRAIN ML 308-
FT 225).
FT VARIANT 845 847 QGT -> LGA (IN STRAIN ML 308-225).
FT VARIANT 855 855 S -> T (IN STRAIN ML 308-225).
FT VARIANT 888 888 Q -> L (IN STRAIN ML 308-225).
FT VARIANT 1025 1025 S -> I (IN STRAIN ML 308-225).
FT CONFLICT 61 63 ETV -> TTT (IN REF 5).
SQ SEQUENCE 1039 AA; 106841 MW; 5170D647C8DEBE0 CRC64;
```

Query Match 4.5%; Score 216; DB 1; Length 1039;

```
Best Local Similarity 21.4%; Pred. No. 0.00012;
Matches 199; Conservative 126; Mismatches 344; Indels 260; Gaps 47;

QY 8 LVLSSFLACFTSCSTVFATAENIGPSDFDGSNT-----GTYTPKNTTGTIDY-- 57
Db 36 VAVALSAAVTSPLVLAADIV--VHPGETVNGGTLANHDQIVFGTNGMTISTGLEYP 93
QY 58 ---TLTGDTITLNLGDSALT-----KGCFSDDTESLSPAGKYSLSFLNLIKSS 103
Db 94 DNEANTGGQWQVGGTANKT---VTSGLQVRNFGGVSVDVVIS---AGGQSLOGRAVNT 150
QY 104 AEG---AALSVTTDKNLSLTGSLTFLAAPSIVTTPSGKG----- 142
Db 151 LNGGEQWHEGATATCTVINDKGQWVKPGTV---ATDTVNTGAEGGPDANGDTGQF 206
QY 143 -----AVKCGDGLT-----FDNG----- 156
Db 207 VRGDVAVRTINKNGROIVRAEGTANTVYVYAGDQVPHGHALDTTLNGGYQYVHNGTAS 266
QY 157 -TILFKODY-CEENGGAISTKNSLKNSTGTSFEGNKSATGKKGGAICATGTVDITNN 214
Db 267 DIVVNSDGHQIVKNGGVAGTATVVKQGR-L-QVDAGGTATNVLKQGGALVTSTAATVGI 325
QY 215 TAPTLFS-----NNTAEAGGAINSTGCTTGNTS-----LVFSNSVTATA-- 257
Db 326 NRLGAFSVVEGRADNVLENGRLD-----VLTGHTATNTRVDDGGTLDVRNGGTATTVS 380
QY 258 -GNGGALSADADVTISGNOS---VTFSGNQAVANGAIYAKK---LTLASG----- 301
Db 381 MNGGVLLADSGAAVSGTSGDKAFSIGGQA-----DALMLEKGSFTLNAGDATTDTTV 436
QY 302 GGGGIFSNNIVQGTAGNGGASILAA-----GPCSLSAEAGDITFNGNALIV 349
Db 437 NGGLTFARGTLAGTTLNNGAILTLSGTVANNITIREGDALL--QGSITGNSVEK 494
QY 350 ATPQOTKNSIDIGSKATITNLRALSIGHSIFYDPDITANTAADSTTLNLNKADAGNST 409
Db 495 SSGSGTLTVSNTT---LTQKAVNLNE-----GTLTLDNSTVTVDVI---AQRGTAL 538
QY 410 DYSGSIVFESGE-----KLSE-----DEA---KVADNL-----TSTLK---QPV 442
Db 539 KITGSTVLNGAIDPTNVTULASATNIPDNATVQSVDDLSHAGQIHFTSTRTGKFPVAT 598
QY 443 LTAGNLVLKRG-VTLDTK-GFTQTAGSSVIMDAG-----TTLKASTEEVTLTGLSIPVDS 495
Db 599 LKVKNLNGQGTISLRVPRDMAQNNADRLVIDGGRATGKTILNLVNGNSASGLA---T 654
QY 496 LGEKKVVTAAASAASKNVALSGPILLLDNQNAYE-----NHDLGKTQDFSFVQLSALGTA 551
Db 655 SKGQIQVBEAINGATTE---EGAFV---QGNRLQAGAFNYSLNRDSDSWSYLRSE--NA 705
QY 552 TTTDVPAPVTPATPHYGYQGTWGTWDDTASTPTKKTATLAWNT-----GYLPNP 604
Db 706 YRAEVP-----YASMLTQMDYDRIVAGSRHSQTGVNGENNSVRLSQGGHLGHD 756
QY 605 ERQG---PLVPNSLMGFSFDIAQGVIERALTLCSDRG---FWAAGVANFLDKDKKGEK 659
Db 757 NNGGIARGATPESS-GSYGFVR-LEGDLMRTEVAGSVTAGYGAAGHSSVDVKDDGSR 814
QY 660 RYRHKSGGYAIGGAQAQTCSENLISFAFQLFGSKDFLVAKNHTDITYAGYIQHITEC 719
Db 815 ACTVRDDAG-CLGGVYLNVLVHTSSGLWADIVAQGRHSMKSSDNDNDFARGW----- 865
QY 720 SGFICGLLDKLPKGSWSHKLPLVLEGOLAYS 748
Db 866 -CWLGSLFGLPFSITDN-LMLEPOLQY 892

RESULT 15
FHAB BORPE
ID FHAB BORPE STANDARD; PRT; 3591 AA.
AC P12255;
DT 01-OCT-1989 (rel. 12, Created)
```

01-FEB-1996 (Rel. 33, Last sequence update)  
01-FEB-1996 (Rel. 33, Last annotation update)  
FILAMENTOUS HEMAGGLUTININ.  
FHAB.  
Bordetella pertussis.  
Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
Bordetella.  
NCBI\_TaxID=520;  
[1]  
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
MEDLINE-90355839; PubMed-2388559;  
Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;  
"Genetic characterization of Bordetella pertussis filamentous  
haemagglutinin: a protein processed from an unusually large  
precursor";  
Mol. Microbiol. 4:787-800(1990).  
[2]  
SEQUENCE OF 1-3261 FROM N.A.  
MEDLINE-89202384; PubMed-2539596;  
Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;  
"Filamentous hemagglutinin of Bordetella pertussis: nucleotide  
sequence and crucial role in adherence";  
Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).  
CC -!- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND  
INFECTION.  
CC -!- SUBCELLULAR LOCATION: SURFACE.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; M60351; AAA22974.1; -  
DR EMBL; M60351; AAA22975.1; ALT\_INIT.  
DR EMBL; M60351; AAA22976.1; ALT\_INIT.  
KW Antigen; Hemagglutinin.  
SQ SEQUENCE 3591 AA; 367420 MW; EF7418B30D6E5138 CRC64;  
  
Query Match 4.5%; Score 215; DB 1; Length 3591;  
Best Local Similarity 21.6%; Pred. No. 0.0007;  
Matches 181; Conservative 86; Mismatches 285; Indels 286; Gaps 41;  
  
QY 55 IDYTLTGDTITLNLGDSA-----ALTKGCFSDTTESL----- 86  
Db 952 VDVNGTGDVVRVAKLYSDAGADLQAGRSMTLG-IVPTTGDQLARAQOKLGLSVKSDGGLQ 1010  
QY 87 SPAGKYSLSFLNLIKSSAEGAALSVTTDK-----NLSLTGFSSTFLAA----- 130  
Db 1011 AAAGALSAAAEVAGALELSGGVTVDRASARARIDSTGSGVIGALKAGAVEAASPR 1070  
QY 131 -----PSSVITTPSGKGVKCG-----GDLTFD-NGTTLFKODYCEENG 169  
Db 1071 ARRALLRQDFFTPGVSVVVRAQGNVTGVRGDPHQGVLAQGDIIIMDAKGGTLLRLNDALTENG 1130  
QY 170 G-AISTKNSLKNSTGTSFEGNKSATGKKG-----GAICATGTVDTTNTAPT 219  
Db 1131 TVTISADSVALEHSTIESKISQSIVLAAGDKGKPAVSKVAKKFLNGTLRAVNDNNETM 1190  
QY 220 FSNNT-----AEAA-----GGAI-----N 233  
Db 1191 SCRDIDVDGRPOITDAVTGEARKDESIVVSDAALVADGGPIVVEAGELVSHAGGICNGRN 1250  
QY 234 STGCTITGNT--SLVFSNSVTATAGN-----GGALSGDADVTISGNQSV- 277  
Db 1251 KENGASVTVRTTGNLV--NKGYSAGKGVLEVGGALTNELVGVSDGTQRTQEAORIENR 1307  
QY 278 -TFSGNQAVANGAIYAK-----KLTASGGGGGGSFNSNNIVQGTAGN- 320  
Db 1308 GTFQSOAPAGTAGLVVVKAAEAIVHDGVNATKGMQIAGKGGSGPTVTAG-AKATTSANK 1366

Qy 321 -----GGAISILAA-----GECSLSAEAGDITFNGNAIVATTPOTTK 357  
Db 1367 LSVDVASHDNAGSLDIKGGQAVVAGRYAEHGEVSIQ---GDYTVSADAI-ALAAQVTQ 1422  
Qy 358 RNSIDIGSTAKITNLRAISGHSIFFYD-----PITANTAADSTDTLNL-----NKA 403  
Db 1423 R-----GGAANLT-----SRHDTREFSNKRIRLMGPLQVNAGGPPSVNTGNLKVREGVTVTAA 1472  
Qy 404 DAGNSTD-----YSGSIVFSGEKLSEDEAKVADNLT---STLKOPVTLTACNVLVLRGVT 455  
Db 1473 SFDNETGAENVAKSATLTTSGAARNAGKMQVKEAATIVAAASVSNPGTETAG-----KDI 1527  
Qy 456 LDTKGFTOTAGSSVMDAGTTLKASTE-----VTITGLSIPVDSLGECK 501  
Db 1528 VTSRGGFNECK---MESNKKDIVIKTEQFSNGRVLDAKHDLTVTASGOADNRGSLKAGHD 1584  
Qy 502 VIAAS-----AASKNVALSGPILLDDNOGNAYENHDL-----GKTQDFSFVQLS-- 546  
Db 1585 FTVQQRIDNSGTMAAGHDATLKAP--HLRNTGQVVAGHDIIINSKLENTGRVYDARND 1642  
Qy 547 -ALGTATTTDVPVPTVATPHYGYQGTGWTWVDDTASTP-----KTKTATLAWTN 597  
Db 1643 IALDVADFTNTGSLYAEHDATLTLAQGTQDRLVVDQDHILEPVAEGTLRVKAKSLTTE-IE 1701  
Qy 598 TGYLPNPEROGPLVPNSLWGSFSDIQATQGVIERALSALTCSDRG-----FWAAG 646  
Db 1702 TG---NP---GSLIAEVO-----ENIDNKAIVGKDLTSSAHGNVANEANALLWAAG 1749

Search completed: October 2, 2001, 03:32:05  
Job time: 240 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2001, 03:30:50 ; Search time 119.05 Seconds  
(without alignments)  
1031.324 Million cell updates/sec

Title: US-09-446-677B-4

Perfect score: 4795

Sequence: 1 MKSFSMLVLSSTLACFTSC.....FEVRGSSRIYVWDGKRFQF 928

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 425026 seqs, 133305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_16:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.oranalle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp rodent:\*  
12: sp.unclassified:\*  
13: sp.vertebrate:\*  
14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4795	100.0	928	2	Q9RB65 chlamydia p
2	4721	98.5	914	2	Q86163 chlamydia p
3	2039	42.5	926	2	P71135 chlamydia p
4	2033	42.4	930	2	Q92393 chlamydia p
5	2028	42.3	930	2	Q9RB66 chlamydia p
6	1924	40.1	928	2	Q92398 chlamydia p
7	1855	38.7	936	2	Q92898 chlamydia p
8	1854	38.7	936	2	Q9J542 chlamydia p
9	1827.5	38.1	772	2	Q9RB71 chlamydia p
10	1786.5	37.3	839	2	P77792 chlamydia p
11	1785	37.2	928	2	Q86164 chlamydia p
12	1785	37.2	949	2	Q9K299 chlamydia p
13	1745.5	36.4	847	2	P71132 chlamydia p
14	1701	35.5	846	2	P71133 chlamydia p
15	1406.5	29.3	841	2	Q923A1 chlamydia p
16	1403	29.3	922	2	Q929G5 chlamydia p
17	1402	29.2	922	2	Q9K1Y9 chlamydia p
18	1393	29.1	922	2	Q924H9 chlamydia p
19	1347.5	28.1	973	2	Q92896 chlamydia p

20	1347.5	28.1	995	2	Q9K2A1 chlamydia p
21	1273	26.5	1276	2	Q9JRW2 chlamydia p
22	1218.5	25.4	1407	2	Q92899 chlamydia p
23	1109.5	23.1	1013	2	Q84879 chlamydia t
24	1106.5	23.1	712	2	Q9RB73 chlamydia p
25	1088	22.7	445	2	Q9RB67 chlamydia p
26	1035	21.6	987	2	Q9PL45 chlamydia m
27	1009.5	21.1	359	2	Q9J5K6 chlamydia m
28	905.5	18.9	867	2	Q9PL41 chlamydia m
29	886	18.5	878	2	Q84882 chlamydia t
30	880	18.4	649	2	P71134 chlamydia t
31	816.5	17.0	494	2	Q9RB68 chlamydia p
32	814	17.0	427	2	Q9RB70 chlamydia p
33	716.5	14.9	978	2	Q92895 chlamydia p
34	711.5	14.8	978	2	Q9RB63 chlamydia p
35	698.5	14.6	1723	2	Q92812 chlamydia p
36	698.5	14.6	1732	2	Q9RB59 chlamydia p
37	698.5	14.6	1732	2	Q9RB59 chlamydia p
38	693.5	14.5	947	2	Q92813 chlamydia p
39	692.5	14.4	39	2	Q9J5E2 chlamydia p
40	682	14.2	1672	2	Q9PY22 chlamydia m
41	679	14.2	946	2	Q92880 chlamydia p
42	679	14.2	946	2	Q92880 chlamydia p
43	654.5	13.6	983	2	Q9PL44 chlamydia m
44	648.5	13.5	1751	2	Q84418 chlamydia t
45	647.5	13.5	1016	2	Q84880 chlamydia t

#### ALIGNMENTS

RESULT 1  
Q9RB65 ID Q9RB65 PRELIMINARY; PRT; 928 AA.  
AC Q9RB65; Q9RB64; Q9S6P2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (OUTER MEMBRANE PROTEIN 5  
DE PRECURSOR) (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).  
GN PMP\_10 OR OMP5 OR CP0303.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CML029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VR1310;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,  
RA Christiansen G., Birke Lund S.;  
RT "Chlamydia proteins containing the GAI-repeat belong to a subfamily  
RT of autocatalytic pathogenicity factors.";  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,  
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
RA Bowman C., Dodson R., Gwynn M., Nelson W., DeBoy R., Kolonay J.,  
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MOpn and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).



DR EMBL: AP002546; BAA98657.1; -  
DR EMBL: AJ133034; CAB37071.1; -  
DR EMBL: AE002192; AAF38160.1; -  
DR TIGR: CP0303; -  
DR InterPro: IPR003357; -  
DR Pfam: PF02385; OMP; 1.  
KW Signal.  
SQ SEQUENCE 928 AA; 97229 MW; 0590D5206A1D0E1 CRC64;

Query Match 100.0%; Score 4795; DB 2; Length 928;  
Best Local Similarity 100.0%; Pred. No. 9, 7e-266;  
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKSQFSLVLSSTLACTSCSTVFAATAENIGPSDSFGSTNGTYPKMTTGGIDYTLT 60  
DB 1 MKSQFSLVLSSTLACTSCSTVFAATAENIGPSDSFGSTNGTYPKMTTGGIDYTLT 60  
OY 61 GDTTLQMLGDSALTKGCFSDTETESLFAKGKYSLSFLNTKSSAEGAALSVTTDKNLST 120  
DB 61 GDTTLQMLGDSALTKGCFSDTETESLFAKGKYSLSFLNTKSSAEGAALSVTTDKNLST 120  
OY 121 GFSLSLFLAAPSSVITTPSGKAVKCGGDLTFPNNGTILFKODYCEBNGAISTKNLSLK 180  
DB 121 GFSLSLFLAAPSSVITTPSGKAVKCGGDLTFPNNGTILFKODYCEBNGAISTKNLSLK 180  
OY 181 NSTGSISEBCKSSATGKKGCAICATGTVDTNTNTAPTLFSNNIAEAGAINSTGCTI 240  
DB 181 NSTGSISEBCKSSATGKKGCAICATGTVDTNTNTAPTLFSNNIAEAGAINSTGCTI 240  
OY 241 TGNLSLSESNVATATAGNGALSGADVTISGNOSVTFSGNOAVANGAIIYAKKILTLAS 300  
DB 241 TGNLSLSESNVATATAGNGALSGADVTISGNOSVTFSGNOAVANGAIIYAKKILTLAS 300  
OY 301 GGGGGISFSNNIYOGTTAGNGAISTILAGECISLAEGDITFNGNAIVATTPQTKRNS 360  
DB 301 GGGGGISFSNNIYOGTTAGNGAISTILAGECISLAEGDITFNGNAIVATTPQTKRNS 360  
OY 361 IDIGSTAKITNLAISGHSIFFYDPTITANTADSTDTLNKNKADAGNSTYSGSIYFSGE 420  
DB 361 IDIGSTAKITNLAISGHSIFFYDPTITANTADSTDTLNKNKADAGNSTYSGSIYFSGE 420  
OY 421 KLSDEAKVADNLTSTLKQPVTLTLAGNLVLRKGYTLDTKFTQTAGSSVIMDAGTTLKAS 480  
DB 421 KLSDEAKVADNLTSTLKQPVTLTLAGNLVLRKGYTLDTKFTQTAGSSVIMDAGTTLKAS 480  
OY 481 TEEVTLGLSLIPVDSLGEKKVYIAASAANKVALSGPIILLNQGAYENHDLGKTQDF 540  
DB 481 TEEVTLGLSLIPVDSLGEKKVYIAASAANKVALSGPIILLNQGAYENHDLGKTQDF 540  
OY 541 SFVQLSALGATTTDPAVPTVATPTHYGYGTGWTGMDVDTASTPKTKATTLAMTNGY 600  
DB 541 SFVQLSALGATTTDPAVPTVATPTHYGYGTGWTGMDVDTASTPKTKATTLAMTNGY 600  
OY 601 LMPEROGPIVNSLWGSFSDIOAIOGVIERSAITLCSDRGFMAAGVAFNLDDKKGEKR 660  
DB 601 LMPEROGPIVNSLWGSFSDIOAIOGVIERSAITLCSDRGFMAAGVAFNLDDKKGEKR 660  
OY 661 KYRHKSGVYIIGAAGQCSENLISFAFCQLGSDKDFLVAKNHTDYAGAFTYIOHTIECS 720  
DB 661 KYRHKSGVYIIGAAGQCSENLISFAFCQLGSDKDFLVAKNHTDYAGAFTYIOHTIECS 720  
OY 721 GFPGCLLDKPLGSMHSKPLVLEGOALVSHVSNLKTYYTAPPEKSGWGNNAFMMMGAS 780  
DB 721 GFPGCLLDKPLGSMHSKPLVLEGOALVSHVSNLKTYYTAPPEKSGWGNNAFMMMGAS 780  
OY 781 SHSPEYLHCFDYPYPIKLNLYIIRDSFSEKTEGSRFSDSNLFLNLPJIGVKEFEKS 840  
DB 781 SHSPEYLHCFDYPYPIKLNLYIIRDSFSEKTEGSRFSDSNLFLNLPJIGVKEFEKS 840  
OY 841 DCNDFSYDLTLSTVPPDLIRNDPKCTALVYSGASWETIYANNLARQALQVRAAGSHYAFSPM 900  
DB 841 DCNDFSYDLTLSTVPPDLIRNDPKCTALVYSGASWETIYANNLARQALQVRAAGSHYAFSPM 900

OY 901 FEVLGQFVEFEVRSRRYVNDLGKRFQF 928  
DB 901 FEVLGQFVEFEVRSRRYVNDLGKRFQF 928

RESULT 2  
086163 PRELIMINARY; PRT; 914 AA.  
ID 086163  
AC 086163  
DT 01-NOV-1998 (TREMblrel. 08, Created)  
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
DE OUTER MEMBRANE PROTEIN 5 PRECURSOR (FRAGMENT).  
GN OMP5.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiales.  
OX NCB1\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPRAIN-CDC/CWL-029/VR-1310;  
RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;  
RT "Identification of two novel genes encoding outer membrane complex  
associated surface layer proteins in Chlamydia pneumoniae."  
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ001311; CA04671.1; -  
DR InterPro: IPR003357; -  
DR Pfam: PF02385; OMP; 1.  
DR Pfam: PF02415; DUF145; 1.  
KW Signal.  
FT SIGNAL. 1 15 POTENTIAL.  
FT CHAIN 16 >914 OUTER MEMBRANE PROTEIN 5.  
FT NON\_TER 914 914  
SQ SEQUENCE 914 AA; 95603 MW; 8BF33B4B680F5E3 CRC64;

Query Match 98.5%; Score 4721; DB 2; Length 914;  
Best Local Similarity 100.0%; Pred. No. 1, 6e-261;  
Matches 914; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKSQFSLVLSSTLACTSCSTVFAATAENIGPSDSFGSTNGTYPKMTTGGIDYTLT 60  
DB 1 MKSQFSLVLSSTLACTSCSTVFAATAENIGPSDSFGSTNGTYPKMTTGGIDYTLT 60  
OY 61 GDTTLQMLGDSALTKGCFSDTETESLFAKGKYSLSFLNTKSSAEGAALSVTTDKNLST 120  
DB 61 GDTTLQMLGDSALTKGCFSDTETESLFAKGKYSLSFLNTKSSAEGAALSVTTDKNLST 120  
OY 121 GFSLSLFLAAPSSVITTPSGKAVKCGGDLTFPNNGTILFKODYCEBNGAISTKNLSLK 180  
DB 121 GFSLSLFLAAPSSVITTPSGKAVKCGGDLTFPNNGTILFKODYCEBNGAISTKNLSLK 180  
OY 181 NSTGSISEBCKSSATGKKGCAICATGTVDTNTNTAPTLFSNNIAEAGAINSTGCTI 240  
DB 181 NSTGSISEBCKSSATGKKGCAICATGTVDTNTNTAPTLFSNNIAEAGAINSTGCTI 240  
OY 241 TGNLSLSESNVATATAGNGALSGADVTISGNOSVTFSGNOAVANGAIIYAKKILTLAS 300  
DB 241 TGNLSLSESNVATATAGNGALSGADVTISGNOSVTFSGNOAVANGAIIYAKKILTLAS 300  
OY 301 GGGGGISFSNNIYOGTTAGNGAISTILAGECISLAEGDITFNGNAIVATTPQTKRNS 360  
DB 301 GGGGGISFSNNIYOGTTAGNGAISTILAGECISLAEGDITFNGNAIVATTPQTKRNS 360  
OY 361 IDIGSTAKITNLAISGHSIFFYDPTITANTADSTDTLNKNKADAGNSTYSGSIYFSGE 420  
DB 361 IDIGSTAKITNLAISGHSIFFYDPTITANTADSTDTLNKNKADAGNSTYSGSIYFSGE 420  
OY 421 KLSDEAKVADNLTSTLKQPVTLTLAGNLVLRKGYTLDTKFTQTAGSSVIMDAGTTLKAS 480  
DB 421 KLSDEAKVADNLTSTLKQPVTLTLAGNLVLRKGYTLDTKFTQTAGSSVIMDAGTTLKAS 480

```
QY 481 TEEVTLGSLSPVSLGSGKKVIAAASAKNVALSGPILLDNOGNAYENHDLGKTQDF 540
DB 481 TEEVTLGSLSPVSLGSGKKVIAAASAKNVALSGPILLDNOGNAYENHDLGKTQDF 540
QY 541 SFVQLSALGATTTDPAVPATPTATPHYGYOGTGMWVDSTAPKTKATLAWMTNGY 600
DB 541 SFVQLSALGATTTDPAVPATPTATPHYGYOGTGMWVDSTAPKTKATLAWMTNGY 600
QY 601 LPNERGGLPVNLSMGFSFDIOAGIYIERSALITLCSDRGFMAAGVAMFLDKDKCKEKR 660
DB 601 LPNERGGLPVNLSMGFSFDIOAGIYIERSALITLCSDRGFMAAGVAMFLDKDKCKEKR 660
QY 661 KYRKSAGYAIAGAAOCSENLISFACCOLFGSDKDLVAKNHDTYAGAFYIOHTICS 720
DB 661 KYRKSAGYAIAGAAOCSENLISFACCOLFGSDKDLVAKNHDTYAGAFYIOHTICS 720
QY 721 GFICLLDLKLPMSHMRPLVLEGLAYSHVNDLKTATYAPYEVKSGMGNAMFMMLGAS 780
DB 721 GFICLLDLKLPMSHMRPLVLEGLAYSHVNDLKTATYAPYEVKSGMGNAMFMMLGAS 780
QY 781 SHSPELHGFDTYAPYIKLNTLYIRDSFSEKTEGSRFSDSLFNLSPIGYKFKFS 840
DB 781 SHSPELHGFDTYAPYIKLNTLYIRDSFSEKTEGSRFSDSLFNLSPIGYKFKFS 840
QY 841 DCNFSYDLTSLYPDLIRNDPKCTALVYSGASMEFYANNLAQALQVRAGSHYAFSPM 900
DB 841 DCNFSYDLTSLYPDLIRNDPKCTALVYSGASMEFYANNLAQALQVRAGSHYAFSPM 900
QY 901 FEVLGQFVFEVRGS 914
DB 901 FEVLGQFVFEVRGS 914
RESULT 3
ID P71135 PRELIMINARY; PRT: 926 AA.
AC P71135;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
DE Putative 98 kDa outer membrane protein.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ABORTION S26/3;
RL Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
  Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U72499; AAB18188.1; -
DR InterPro: IPR003357; -
DR Pfam: PF02385; OMP; 1.
DR Pfam: PF02415; DUF145; 1.
SQ SEQUENCE 926 AA: 98439 MW: 3E755E2F594750F CRC64;
```

Query Match 42.5%; Score 2039; DB 2; Length 926;  
Best Local Similarity 46.2%; Pred. No. 1,9e-108;  
Matches 442; Conservative 137; Mismatches 318; Indels 60; Gaps 18;

```
QY 1 MKSQFSLVLSSTLAC-----FTSCSTVPAATLAEINICPSDFGSGTGTGTTAKNTT-- 53
DB 1 MRSLYKILSLSTLTLPISFHSOLAHEVALTQESI-----LDA--GAFSPSTSTAG 52
QY 54 GIDTTLTGDTLQNLGSAALTKGCFSDTRESLFAKGYSFLNLTSSAEGALSV-T 112
DB 54 GTIYNVSDISIDYVCGTALALASAFVQTADNLTFFKGNHSLSTTANAAANRAGIVNT 112
QY 113 TDKNLSLTGFSSTLTAAPSSVITTPSGAGVACGGLTDPNNGTTLFKDYCEENGAI 172
DB 113 ADKILTLTDFSKLSFKKCPSSLVNT--GKAMKSGALANLANNASILFPDNTSANGAI 170
```

```
QY 173 STKSLSKNSNGSTSPFGKNSATKKGKGAICATGVTDTNNAPTLFSPNNIAEAGAI 232
DB 171 SCRASFSLGSSKELSFETT--STAKKGALAAATGIALHLSNOCTIRFSGTAAVNSGAV 227
QY 233 NSTGNCITGNTSLFSENSVTATA--GNGAL--SGDA-DVTISGNSVTFSGNOVA 286
DB 228 YSEASMTIAGNNHAFENNNVSSSDCGGAIHCSKTSANLTLIRNKNVLIFFENNTSSA 287
QY 287 NGCAIYAKKLTLASGGGGISFNNIYQTTAGNGAISIIAGCSLSAEGDITNGN 346
DB 288 KGAIFYDKLITL--GCPFAFINNKVTHAP--KGAIGIAGANECSTLAEHGDITFDNN 344
QY 347 AIVATTFQTRKRNIDIGSTAKITNLASGHSIFEDOPTAANAASTDLNINKMDAG 406
DB 345 LMATQDNATIKRNAINIEGNGEYVNLRAASGKTSIFDTPIVE--GNADLLINKKE-G 401
QY 407 NSTDSSGSIYFSGEKLSEDEAKVADNLTSTLQKQVTLTACNLVLRGVTLDTKGFTOTAG 466
DB 402 DKT-YNRIIFSGEKLTEBOAAVADNLTFTPTQITLAGEVLRSQVEYAKTVQOTAG 460
QY 467 SSVIMADGTTLKASTEYVTLGLSIPVDSLGEKKVIAAASAKNVALSGPILLDNOG 526
DB 461 SLIIMDAGTKLSAKTEADATITNLAINPNTLDGKKFAVDAVAAKKNVTLSGAIGVIDPTG 520
QY 527 NAYENHDLGKTQDFSFVQLSALGATTTDPAVPATPTATPHYGYOGTGMWVDSTAP 586
DB 521 KEYENHDLNDTLALGIGQLSGKSVTTTNVSHVGYAETHYQGMWVSWVDONSDP 580
QY 587 KTKATLAWNTGYLIPNERGGLPVNLSMGFSFDIOAGIYIERSALITLCSDRGFMAA 645
DB 581 KTOFAITWNTKGTGVNPERARPLVNLMSGFTLDSIODVLEKRSVDSILETRGLMVS 640
QY 646 GVANFLDKKKGKERRKHSRSGAIGAAOTCENLISFACCOLFGSDKDLVAKNHTD 705
DB 641 GIGNFFKDRNAENRKRHSISGVLGATNTSRSDLSVAFCOLFKDKDYLSKNAAN 700
QY 706 TYGAFYIOHTE-----CSGFICGLDLKLPMSHMRPLVLEGLAYSHV 751
DB 701 VYASVYIOHVSFKDILTRLENGPNTCCSGFS-----KEIFLDAQITTYCHTA 749
QY 752 NDLTKTATYAPYEVKSGMGNAMFMMLGASSHYPEYLHCFDYAPYIKLNTLYIRDSFS 811
DB 750 NNMVTSYTDPEVAGSGMGNPDLGLTSTVPIPVSSSIFDSVAPFAKIQVVAHQDOK 809
QY 812 EKGTEGSRFSDSLFNLSPIGYKFKFSDCNFSYDLTSLYPDLIRNDPKCTALVY 871
DB 810 EPTTEGAVFESSDILANVSPIGIFKELISYGERSAVDLTMYIPDYVRHNSCMTGLAIN 869
QY 872 GASMEFYANNLAQALQVRAGSHYAFSPMEYVLGQFVFEVRGSRIYNDIGKRFQF 928
DB 870 DSVLTTATNLARQALTVRAGNHTALTSGVEMFSQGFELKSSSRNRYNDIGAKVAF 926
RESULT 4
ID Q9Z393 PRELIMINARY; PRT: 930 AA.
AC Q9Z393;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
DE OUTER MEMBRANE PROTEIN 11 PRECURSOR.
GN OMP11 OR PMP-8.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
  Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
  Comparative genomes of Chlamydia pneumoniae and C. trachomatis.;
  Nat. Genet. 21:385-389(1999).
```

RN [2]  
 SEQUENCE FROM N.A.  
 RC STRAIN=VR1310;  
 RA Hierro K., Boesen T., Daugaard L., Knudsen K., Madsen A.,  
 RA Christiansen G., Birkelund S.,  
 RT "Chlamydia proteins containing the GGAI-repeat belong to a subfamily  
 RT of autoantigenic pathogenicity factors."  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE001627; AAD18590.1;  
 DR EMBL: AJ133034; CAB37068.1;  
 DR InterPro: IPR003357;  
 DR InterPro: IPR003366;  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.  
 DR Signal.  
 FT SIGNAL  
 FT CHAIN 1 27 POTENTIAL.  
 FT CHAIN 28 930 OUTER MEMBRANE PROTEIN 11.  
 SQ SEQUENCE 930 AA; 97669 MW; 46A9B5E3B913C4C CRC64;

Query Match 42.4%; Score 2033; DB 2; Length 930;  
 Best Local Similarity 44.6%; Pred. No. 4,3e-108;  
 Matches 424; Conservative 158; Mismatches 325; Indels 44; Gaps 17;

QY 1 MKSOFSLVLSSTLACTSCSTVFATAENIGPSDFDSTNGTYPKNT--TTGIDYT 58  
 DB 1 MKIPLHLKLLISLTVPILISITAYGADASLSPDSDGAGS-TTPKSTADANGTNYV 59  
 QY 59 LTGDITLONIGDASALTKGCFSDTESLSPAGKGYSLFLINKS-SAEGAALSTYTDKNL 117  
 DB 60 LSGNVTINDAGKGTALGCCFETTGDTFTGKGYSFSTVVDGNSAGAAASTADKAL 119  
 QY 118 SLTGFSLLFLAAPSIVTTPSGKAVCGGDLTFDNNGTILFKODYCE--NGAIST 174  
 DB 120 TFGFSNLSFLAAPT--TVASGKSTLSAGALNLTONGTILFQSNVNEANNNGAIT 177  
 QY 175 KNLSTKSTGSISEFGKSSATGKGAICATGTVDTNTTAPLFSNNIAEAAGALNS 234  
 DB 178 KTLISIGNTSITFTFSNSAK--KLGAISYSAASISGNTGQLVFNNNGETGGALGF 234  
 QY 235 TGNCTGTGNTSLVSESVATAGNGAL---SGDA-DVTISGNSGTSGNOAVANGG 289  
 DB 235 EASSITONSLSPFGSGTATDAAGKGAICETGEPITLTKNSKSLTRENSSVTOGG 294  
 QY 290 AYAARKLLTASGGGGISFSNNIVQGTAGNGAISIILAECSLSAEGDITFNGNAIV 349  
 DB 295 AICAHGDLISA--AGPILFSNNRCGNTAAGGAIADSGSISLSANOGDITFLGNTLT 352  
 QY 350 ATPTQTKRRSIDGSAKTITNLRAISGHISFTFDPITTAADSTTLNLRADAGNST 409  
 DB 353 STSAPTSTRAIYLGSSAKITNLRAAGOSIYFDPITASMTG-ASDYLITINPDNSPL 411  
 QY 410 DYSGSIYFSEKLSSEDAKADNLSTLKQPVTLTLAGNLVKRCVLTDTGFTGTASSV 469  
 DB 412 DYSGTIVFSEKLSADAKADNLTSLKOPALASSTLAKGVELDVAGFYOTBESTL 471  
 QY 470 IMDAGTLAKASTEETVLTGSLIPVDSLEGKKVYIAASAASKNALSGPILLDNOGNAY 529  
 DB 472 LMPGFTLKAADTEAISTLTKLVVDLSALEGNKSVIETAGANKTITLSPVFOGSSNFI 531  
 QY 530 ENMDLCTOQFSVVOALSATATTT-----DVPAPVTATP--THYGOGTGWTWDDT 582  
 DB 532 ESTHTIQ---ATQPLVLTATTAASDIYIDLTLSPVOTPEPHYGOGHWEATWMD-- 585  
 QY 583 ASPTKKTATLATNTGYLNPPEOGPLVPNSLMSGFSFOALQGVGERALTLCSDRGF 642  
 DB 586 --ISTAKSGMTVTTCYNNPERRASVVDLSLMASTGTDITRTIQOIMTSQANSITQOAKL 643  
 QY 643 WAAGVANFLDKKGGKRRYRHSGGVIGAAQCSENLISFAFCQLFGSDKDFLVAKN 702  
 DB 644 WAGGTANFHKDKSGTNOAFRRHSKYGVIGSAEDSENFISVAFCQLFGKDXDLFIVEN 703  
 QY 703 HTDTYGAFTYIOHTECSGF----IGCLLDKLPGSMHKLVLLEGQLAYSHVSNDLTKTY 758

DB 704 TSHNYLASLYQHRAPFLGGLPMPSFGSITDML---KDIPILINAQLSYSTYKNDMDTRY 759  
 QY 759 TAYEVEVSGNGNNAFNMMGLASSHSY--PEYLHCEDTYAPYIKLNTLYIRDDSEKTEG 817  
 DB 760 TSTPEAGSGWTNNSGALGLSLALPLKAPFPQGFPGFLKFOAVYSROONRESGAA 819  
 QY 818 RSPEDSNLFLNLSLPIGVKFEKFSQNDPSYDLTSLVDPDIRNDPKCTALVIGASMET 877  
 DB 820 RAPDGDVLNCSIPVGRLEKISDEKNNNEISLAYIGDYRRKRPRTSLWVSGASWTS 879  
 QY 878 YANNIARQALQVRAGSHYAPSPMEFVLQGVFEYKSGSRINYNDLGKEFP 928  
 DB 880 LCKNIARQAEFLASAGSHLTLSPHVELSGEAYELRGSAAHYNNDCGLRYSF 930

RESULT 5  
 Q9RB66 PRELIMINARY; PRT; 930 AA.

AC Q9RB66;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE  
 DE PROTEIN G FAMILY).  
 GN PMP\_8 OR CP0307.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shihai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kihara S., Nakazawa T.,  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CML029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Ulfback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia  
 RT pneumoniae AR39."  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AE002546; BAA98654.1;  
 DR EMBL: AE002193; AAF38164.1;  
 DR TIGR: CP0307;  
 DR InterPro: IPR003357;  
 DR InterPro: IPR003368;  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.  
 SQ SEQUENCE 930 AA; 97639 MW; 46A8896761391C09 CRC64;

Query Match 42.3%; Score 2028; DB 2; Length 930;  
 Best Local Similarity 44.5%; Pred. No. 8,3e-108;  
 Matches 423; Conservative 158; Mismatches 326; Indels 44; Gaps 17;

QY 1 MKSOFSLVLSSTLACTSCSTVFATAENIGPSDFDSTNGTYPKNT--TTGIDYT 58  
 DB 1 MKIPLHLKLLISLTVPILISITAYGADASLSPDSDGAGS-TTPKSTADANGTNYV 59  
 QY 59 LTGDITLONIGDASALTKGCFSDTESLSPAGKGYSLFLINKS-SAEGAALSTYTDKNL 117  
 DB 60 LSGNVTINDAGKGTALGCCFETTGDTFTGKGYSFSTVVDGNSAGAAASTADKAL 119  
 QY 118 SLTGFSLLFLAAPSIVTTPSGKAVCGGDLTFDNNGTILFKODYCE--NGAIST 174

```

Db      120 TETFSNLSFLIAPGT--TVASGKSTLSAGALNLTGNGTLTFQSNVSNFANNNGAIFTA 177
      175 KNLSTKSTGSISEFGKSSATGKGAICATGIVDTNNAPLPSNNIAEAAGAINS 234
      178 KTLISGNTSITFTTSNAK---KLGAIIYSSAAASISGNTGQVFPNNNGETGGALGF 234
      235 TGNCTTGNTSLVSENSVATAGNGAL----SGDA-DVTISGNSVTFSGNAVANGG 289
      235 EASSSTIQQNSLFFSGTNTAIDAAKGAIYCETGETPTLTISCKSLTIRENSSVTQGG 294
      290 AIIAKKLTLLAGGGGGGSISSNNIVQGTAAAGCAISILAGECSLSAEADITFNGNAIV 349
      295 AICAHGLDLA--AGPTLFSNNCGNTAAAGKGAIAIDAGSLSLSANQGDIFFLGTLT 352
      350 AATPPTTKRNSIDIGTAKITNLRAISGHSIFPYDPTTANTADSTDTLNLKADAGNST 409
      353 STSAPSTRRAIYLGSSAKITTLNRAAGGOSIYFDPIASNTTG-ASDVLITINQPSNPL 411
      410 DYGSIYFSGEKLSEDEKAYADNLTSTLKOPVTLTGNLTKGVTLDITGEGTQAGSSV 469
      412 DYGSIYFSGEKLSEDEKAYADNLTSTLKOPVTLTGNLTKGVTLDITGEGTQAGSSV 471
      470 IMDAGTTLKASTEVEVLTGLSIPVDSIGEGKKVYIAASAANKVNLSGPILLDMQNGAY 529
      472 LMPGRTKLKADTEAISLTKLVVDLSALEGNKSVIETAGANKTITLSPVFDSSGNFY 531
      530 ENHDLGKTQDFSVQSLATATTT-----DYPAYVTVATP--THGYGQGTWGTWYDVT 582
      532 ESHITNG---AFTQPLVYFTAAATAADIIYIDLALNLPVQTPPEPHGYGQHWETATMD-- 585
      583 ASPRTKTATLAWNTGYLNPPEHOGPLVPSNLWGSPSDIOALOGVTERSLATCSDRGF 642
      586 --ISTAKSGMTWTWTGYNPNPERRASVVDLSMASTDTIRTLQOINTSOANSIYQGR 643
      643 WAAGVANFLDKDKKGEKKRYRHKSGGYAIGAAOTCSENLISAFCOLFGSKDKFLVAKN 702
      644 WASGTANFPHKDKSGTNOARHKSYGYVIGSAADEFSENIFFSAFCQLPFKDKDLFIVEN 703
      703 HTDTYAGAFYIQHTTECSGF----IGCLLDKLGSMHKLAVLEGOLAYSHVSDLTKEY 758
      704 TSHNYLASLILQRAFLGLGPMPSFGSITDML---KDPIILINAOISYTRKNDMDTRY 759
      759 TAYPEVKGSGNNAFNMLGASSHSY--PEYLHCFDYAPITKINLTJYIRDSESEKTEG 817
      760 TSTPEAGSGTNNNSGALIEGSLALILPKAPFPQGTFFPLKQAYVSKROONKESGAEA 819
      818 RSFDDSNLFNLSLPIGVKFEKFSDCNDPSYDLTSLVYDPLIRNDPKCTALVIGASMET 877
      820 RAVDDGDLVNCISLPVGRLEKISEDEKNNEISLAYIGDYRRKNRPRERTSLMWSGASWTS 879
      878 YANNLRQAQLQVRAAGSHYAFSPMEVVGQVFEVYKSSRIYNDVGKGFQF 928
      880 LCKNLARQAFLASAGSHLTLSPHVELSGEAAVYELRGSAAHIYNDVCGILRYSF 930

```

## RESULT 6

```

ID      092398      PRELIMINARY:      PRT;      928 AA.
AC      092398;
DT      01-MAY-1999 (TREMblrel, 10, Created)
DT      01-MAY-1999 (TREMblrel, 10, Last sequence update)
DT      01-MAR-2001 (TREMblrel, 16, Last annotation update)
DE      OUTER MEMBRANE PROTEIN PRECURSOR (POLYMORPHIC MEMBRANE PROTEIN G
      FAMILY).
GN      OMP10 OR PMP_9 OR CP0306.
OS      Chlamydia pneumoniae (Chlamydia pneumoniae).
OC      Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX      NCBI_TaxID=83558;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-CWL029;
RX      MEDLINE=9206606; PubMed=10192388;

```

```

RA      Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA      Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT      "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL      Nat. Genet. 21:385-389(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-VRI310;
RA      Hierro K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RT      Christiansen G., Birkelund S.;
RT      "Chlamydia proteins containing the GGAI-repeat belong to a subfamily
RT      of autotransporting pathogenicity factors.";
RL      Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-J138;
RX      MEDLINE=20350349; PubMed=10871362;
RA      Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA      Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT      "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT      from Japan and CWL029 from USA.";
RL      Nucleic Acids Res. 28:2311-2314(2000).
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN-AR39;
RX      MEDLINE=20150255; PubMed=10684935;
RA      Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA      White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
RA      Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA      Bowman G., Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J.,
RA      McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT      "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT      pneumoniae AR39.";
RL      Nucleic Acids Res. 28:1397-1406(2000).
DR      EMBL; AE001628; AAD18591.1; -.
DR      EMBL; AUI33034; CAB37069.1; -.
DR      EMBL; AP002546; BAA98655.1; -.
DR      EMBL; AE002192; AAF38163.1; -.
DR      TRIM; CP0306; -.
DR      InterPro; IPR003357; -.
DR      InterPro; IPR003368; -.
DR      Pfam; PF02385; OMP; 1.
DR      Pfam; PF02415; DUF145; 1.
KW      Signal.
FT      CHAIN      1      27      POTENTIAL.
FT      SIGNAL     28      928      OUTER MEMBRANE PROTEIN.
FT      CHAIN      28      928
SQ      SEQUENCE 928 AA; 98332 MW; 58910ABF04F12219 CRC64;

```

Query Match 40.1%; Score 1924; DB 2; Length 928;  
 Best Local Similarity 44.5%; Pred. No. 7.1e-102; Indels 36; Gaps 18;  
 Matches 421; Conservative 151; Mismatches 338;

```

      1 MKSQSWLVLSLTACPTSCS-TVFAATAE-NIGPSDFSGSTWGTGTYPRKNTTGID-- 56
      11 KMSLSLHFLISSSLADPLSLNFSFAVVEINLGPNTSFGS---PGYTFPPAQGTNADGT 57
      57 -YLTGDTITLQNLGDSALNKGCFSTTESLSPAGKISLSLNLIKSAGCAALSVTTDK 115
      58 IYNLTGTVSITNAGSPYALASCFKETTGNLSFGHGVOPLNDIADGACFTFNTANK 117
      116 NLSITGSSITFLPAAPSSVITTPSGKAVCGGDLTFDNNGTILFKDYCEENGALSTK 175
      118 LILFSGFSYSLIQTNNAT----TGIGAIKSTGACSIQSNYSCTFGNFSDNDNGALQGS 173
      174 SISL-SLNPMLTFPAKNKAT---QKGAISTGTGITTNNLTNSASFSEBNTAANNGAIYTE 229
      176 NLSLKSTGSISEFGKSSATGKGAICATGIVDTNNAPLPSNNIAEAAGALINST 235
      236 GNCITIGNTSLVSENSVYTAGNGAL---SGDAD---VTISGNSVTFSGNAVANGG 269
      230 ASSFISKNKAIISPTNNSVYTSATGGAIVCSSTAPRPVTLTLDNGNLINFTAITISGG 289
      290 AIIAKKLTLLAGGGGGGSISSNNIVQGTAAAGCAISILAGECSLSAEADITFNGNAIV 349

```

```
Db 290 AYTNDNLVLS--CGPTLFKNNSAIDTAADPGALAIADSGSLSLSALGDDITREGTIVV 347
QY 350 --ATTPTOTNRNSIDIGST-AKTNLRAIGHSTFEEDPTITANAASTDTLNANKDAG 406
Db 348 KGSASSTPTNRNINININAKTIVQRAOGNTYFDPTITSTAALSDALNANGDLA 407
QY 407 NSTDYSGLFSGEKLSEDEAKVADNLSTLKOPVTLTAANLVKRGVTLDTKGFOTAG 466
Db 408 GNRAYOGTIVFSGEKLSEAEADNKSITIQOPLTLAGOLSLKSVTLVAASFSGSPG 467
QY 467 SSVYMDAGTTLKASTEEVTLTGLSIPVDSLEGCKVYIAASAANKVALSGPILLDNOG 526
Db 468 STLMDAGTTLT-TADGITINNVLANDSLKETKATLKATQASQFTVLGSLSLVDPG 526
QY 527 NAYENHDLGTODPSFVQALSGAT--TTDVPAPVPAVPAPTHVGYOGTGMVVDPTAS 584
Db 527 NVEDYSWNNPQVFSCTLTTLADDPANIHITDLAADPLEKNNIHMGYGNALMSQEDTAT 586
QY 585 TPTKTATLTAMTNGYLPNPEROGPLVNSLWGSFSDIOAQVIERATLTCSDRGFWA 644
Db 587 --KSKATTLTWTKGTGYPNPERGRTLVANTLWGSFVDRSTIOQLVATKVKOSQETRGIMC 644
QY 645 AGVAFNLKDKKCKEKKRYRHRKSGGYAIGAAQTCSNLSIFAFQCLFGSDKDLVANKHT 704
Db 645 EGISNFEHKSSTKINKGFRHISAGYVYGATTTLASDNLITAAFCQLGKDRDHPINKRA 704
QY 705 DTYAGAFYIOHIECGSFICGLDKLPDSWSHKPLVLEGOLASHVNSDKTYATAPEV 764
Db 705 SATVAASHLQHLVLS--PSLLKATLPGSEDEPVLEDAOSTYISKNYNTMYTQAPKG 762
QY 765 KGSWGNNAFNMMLGAS-SHSYPEYLHCFDYARYIKLNTYIRDSFSEKTE-GRSFDD 822
Db 763 ESSWYNDGALCELLASSLPHTALSHGELFHAVPPITVEAYIHHQDSKENTLTVRSFDS 822
QY 823 SNFNLSPGLGVKFEKSDCNDSTYDLTSLVDDLIRNDPKCTTALVIGASMETVANNL 882
Db 823 GDIINVSVPGLIFEFERSRERASYTEATVYVADYVKKRNPDCCTTALINNTSWKTGTNL 882
QY 883 AROALQVRAGSHYAFSPMEVIGOFVEVGRSSRYVNDVGGKFOF 928
Db 883 SRQAGIGRAGIFAFSPNLEVTNLSMEIRGSSRYNADLGKFOF 928

RESULT 7
Q92898 PRELIMINARY: PRT: 936 AA.
AC 092898;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE POLYORPHIC OUTER MEMBRANE PROTEIN G FAMILY.
GN PMP. 7.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kallan S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
  Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
  "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
  Nat. Genet. 21:385-389(1999).
RL EMBL: AE001627; AAD18589.1; -.
DR InterPro: IPR003357; -.
DR InterPro: IPR003368; -.
DR Pfam: PF02415; DUF145; 1.
DR Pfam: PF02415; DUF145; 1.
SQ SEQUENCE 936 AA; 100079 MW; 8881D78A53D194EC CRC64;
```

Query Match 38.7%; Score 1855; DB 2; Length 936;

Best Local Similarity 42.0%; Pred. No. 6,3e-98;  
Matches 399; Conservative 172; Mismatches 343; Indels 36; Gaps 18;

```
QY 1 MKGSFNLVLSSTLACTSGSYFAATAENIGSPSDSGSTNGTYPRKMT--TGIDY 57
Db 1 MKSSVNLFFSS-IPLESSLIV-AAEVTLDDSSNNSDGSNGT-TFVFSTTDAAAATTY 57
QY 58 TLNGDITLQMLGD-SAAITGCFSDTEESLFAKGYSL--SFLNKSASGALSVTTD 114
Db 58 SLSDVSFQNALGLITLACGCFLEAGDGLTFQGNQALFAFINAGSSAGYASTAAD 117
QY 115 KNLSTGFSLLFLAAPSYYTTPSGKAVKCGDGLTFDNNNGTILFKODYCEENGAI 174
Db 118 KNLFFNFSRLSITSCP-SLLSPTGQALKSVGNLSLTGNSQIITFQNPSSNGGYINT 176
QY 175 KNLISKSTGSSISPEGKSSATGKKGALCATGTVDTNTNMTAPLIESNNIAEAGA 234
Db 177 KNFLSGTSQFASFSRNO-AFTGKGGVAVATGTTTENSPLGVFSQNALKSGGALYS 235
QY 235 TGNGTITGNTSLVSENSV-TATANGAL--SGADAVTISGNQSVTFSGNOAVANGA 290
Db 236 TDNCSTIDNFOVIFDGSAMEAQAOGALICCTTKTYTLTGKKNLSFTNNATLTGGA 295
QY 291 IYAKKTLASGGGGISFSNNIVGTTA--GNGAISIIAEGCSLSAEGDITFNGAI 348
Db 296 ISGLKVAISA--GGPTLFQSNL-SGSSAGGGGALINIASAGELATASGDTFNNQV 352
QY 349 VATTPTOTNRNSIDIGSTATITNLRAISGHSTFEEDPTITANAASTDTLNANKDAG 408
Db 353 --TNGSTSTNAINIIDTAATVSTIRAAVTSQSIIFYDITPGRASTDNLNKAADNSE 410
QY 409 TDYSGSIVFSGEKLSEDEAKVADNLSTLKOPVTLTAAGNLVLRGYTLIDKGFOTAGSS 468
Db 411 IEKGALVEFGEKLSPEKRIAAANVSTIRQPAVLARGDLVDGYVVFVKDLTQSGSR 470
QY 469 VINDAGTTLKASTEEVTLTGLSIPVDSLEGCKVYIAASAANKVALSGPILLDNOG 528
Db 471 ILMDGGTTLTAKKANLSINLAVNLSDTGNKAALTEADNINISLSTGIALIDEGSF 530
QY 529 YENHDLGTQDFSFVQALSGATTTDVPAPVPA--TPTHVGYOGTGMVVDPTAST 585
Db 531 YENHNLKASATPLLELTITAGANGTITLGLSTLIDPEPTHGYOQNMOLSNANATSS- 589
QY 586 PKTATATLANTNGYLPNPEROGPLVNSLWGSFSDIOAQVIERATLTCSDRGFWA 645
Db 590 --KIGSINMTFRGTGYPSPRKSNNLPLNSLGNFIDIRSIQNLIEFKSGEPPERELMLS 646
QY 646 GYANFLDKKCKEKKRYRHRKSGGYAIGAAQTCSNLSIFAFQCLFGSDKDLVANKHT 705
Db 647 GIANFEYRDMPTRRHGRHISGVALGITTATTPAEDQLTFAFQCLFARDNNHTTKRNGD 706
QY 706 TYAGAFYIOH--ITECSGFICGLDKLPDSWSH---KPLVLEGOLASHVNSDKTY 758
Db 707 TYGASLYFHHTEGFLDIANLWKGKATRAPVLSISITILPSDAKFSYIATDHNKTTY 766
QY 759 TAYPEYKSGWGNNAFNMMLGASSHSYPEYLHCFDYARYIKLNTYIRDSFSEKTEGR 818
Db 767 TDNSIITKGSWRNDAFCADLGASLPVYISVPLYLKEVEPPKVGYYIARHODDFERNAEGR 826
QY 819 SPDSNLNLSLPIGVKFEKSDCNDSTYDLTSLVDDLIRNDPKCTTALVIGASMETV 878
Db 827 AFKSKELINVEIIGVFEEDSKSEKGYDLTLMYILDAARRNPKCOTSLIADANMAY 886
QY 879 ANNLRQALQVRAGSHYAFSPMEVIGOFVEVGRSSRYVNDVGGKFOF 928
Db 887 GTNLARQGESVRAANHQVPMHEITFGQFAFEVRSRRNTNVLGSKFCF 936

RESULT 8
Q9JSA42 PRELIMINARY: PRT: 936 AA.
AC 09JSA42;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
```

```

DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
DE 01-MAR-2001 (TReMBLrel. 16, last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE
DE PROTEIN G FAMILY).
GN PMP_7 OR CP0308.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kubara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RX EMBL: AF002546; BAA98653.1; -
RX EMBL: AE002193; AAF38165.1; -
RX TIGR: CP0308; -
DR InterPro: IPR002637; -
DR InterPro: IPR003357; -
DR InterPro: IPR003368; -
DR Pfam: PF02385; OMP; 1.
DR Pfam: PF02415; DUF145; 1.
DR ProDom: PD004952; -; 1.
SQ SEQUENCE 936 AA; 100105 MW; 3981DB3C950AF95A CRC64;

Query Match 38.7%; Score 1854; DB 2; Length 936;
Best Local Similarity 42.0%; Pred. No. 7.2e-98;
Matches 399; Conservative 172; Mismatches 343; Indels 36; Gaps 18;

QY 1 MKSQFSLVLSLACTSCSTVPATANIIGSDSFGSTNGTTPKNTT--TGIDY 57
DB 1 MKSSVSLPFS--IPLFSSLSIV-AAEVTLDSNNNSYDGSNGT-TFTVFSTDAAGTTY 57
QY 58 TLNGDITLQGLD-SALATKGCFSPTTESLSPKQKSYL--SFLNIKSSADGALSTTTPD 114
DB 58 SLISDVSPONAGALGIDPLASGCFLEAGGDLTFOGNOHALKFAFIMAGSSAGYVASTAAD 117
QY 115 KNLSTLGFSSILFLAAPSIVTTPSGKAVCGGDLTFDNNGTILFKODCEENGAIST 174
DB 118 KNLLENDFSLSLITSCP-SLILSPTGCAKSYGNLSLFTNSQIITFNQSSNDGVIINT 176
QY 175 KNLSTLKNSTGISIFEGNKKSSATSKKGAICATGTVDITNNTPALPFLSNINIAEAGAINS 234
DB 177 KNLFLSGTSQFASFSSRNQ-AFTGKGGGVYVATGITITENSPGIYSFSONLAKSGGALYS 235
QY 235 TGNCTTGNTSLVFSNSV-TATAGNGAL---SGADVTISGNSOSTFEGNOAVANGGA 290
DB 236 TDNCSTIDNKQVIFDGSNAEAAOAGCAICTTDTKTVTLTGKKNLSFTNNALITGGA 295
QY 291 IYAKKILTLASGGGGSIFSNIVOGTTA--GNGAISILLAGCSLSAEAGDITFNGCAI 348
DB 296 ISGLKIVSA--GGPTLFGQNT--SGSSAGGGGGAIIINAGELALSATSGDITFNNNOV 352
QY 349 VAITPOTKRNNSIDIGSTAKITNLRALSGHSIFPDPITANTADSTDTLNLKADAGNS 408
DB 353 --TNGSTSTFNAINIITDAKVTSTIRATGOSIYFYDPITNPGTAASDTLNLNLANSE 410

```

```

QY 409 TDYSGSIVFSGEKLEDEAKVADNLSTLKOPYTLTAGNLVLRKGVTLDTKGFOTAGSS 468
DB 411 IENGGAIVFSGEKLSPPEKAIANAVTITROPAYLANGDLVLRGVYVTKDLTQSGSR 470
QY 469 VINDAGTTLKASTEEVTLTGISPVDSIGREKVVIAASAASKNVALSGPILLDNOGA 528
DB 471 IIMDGCTLSKKEANLSLNGLAVALSLDGTNKAALTEADKNISLSGTALIDTGSF 530
QY 529 YENHDLKQDDEFVQSLAGTATTVDPAVPIYA---TPHYGQGTWGTWDDTAST 585
DB 531 YENHNLKASASTYPLLETLTGAAGCTITLGAISTLLOEPETHYGOENWOLSMANATSS- 589
QY 586 PKRTATLANTNTCYLLENPERGPLYPNLSMGSPSDIOAGVTERALTLCSRGPFAA 645
DB 590 ---KIGSINNTRTGYIPSPERKSNLPNLSMGNTIDRSIQIETSSGPFREIMLS 646
QY 646 GVAFLDKRKGKRRKRRHSGVAIGAQAOTCSENLISFACOLFGSDKDFLAKNHTD 705
DB 647 GIANFPRDSMPTRHGRHISGVALGITATTTPAEDQUTFAFQGLFARDNRHITGKNHGD 706
QY 706 TYGAFYIQH---ITECSGFLGLDLKLPQSWH---KPLVLEGQALYSVNDLTKY 758
DB 707 TYGASLYFHTHTGLFDIANFLMGKATRAPVWLSISQIIPLSFPAKPSYLTDDHMKTY 766
QY 759 TAYPEVKGSMGNNAFNMALGASSHYPEYLHCEPTYAPYIKNLTYLRDSESEKTEGR 818
DB 767 TDNSTIGSKMRNDFCADLGSLEPVIYSVPLLEKEVEPVKQYITVAHQDFTERYAEG 826
QY 819 SFDSNLFNLSLPIGVKFEKSDCNDFSYDLTISYVDLIRNDPKCTTAAVIGASMEY 878
DB 827 AFNKSSELINVEIPGIVFERDSKSEKTYDITLMTIADARRNKQCTSLASDANMAY 886
QY 879 ANNLARQALOVRAGSHTAFSBSMEFVLQGFVEVGRSSRIYNDLGKFOF 928
DB 887 GTNLARQGFVVRANHPQVNPHEIFQFAFEVRSRRYNTNIGSKFCF 936

RESULT 9
Q9RB71 PRELIMINARY; PRT; 772 AA.
ID Q9RB71;
AC Q9RB71;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
DE 01-MAR-2001 (TReMBLrel. 16, last annotation update)
DE PMP_3.
GN PMP_3_2.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kubara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
RX EMBL: AF002545; BAA98226.1; -
RX EMBL: AF002546; BAA98226.1; -
RX InterPro: IPR003357; -
RX InterPro: IPR003368; -
RX Pfam: PF02385; OMP; 1.
RX Pfam: PF02415; DUF145; 1.
SQ SEQUENCE 772 AA; 82931 MW; 484FC56D635801EB CRC64;

Query Match 38.1%; Score 1827.5; DB 2; Length 772;
Best Local Similarity 48.2%; Pred. No. 1.8e-96;
Matches 379; Conservative 117; Mismatches 260; Indels 31; Gaps 11;

QY 158 ILFKODYCEENGAISTKNLSKNSTGISIFEGNKKSSATSKKGAICATGTVDITNNTPAP 217
DB 1 MLPSKNSTONGAITAKTSLTNGTWSALFSEWTSS---KKGAIQTSALTITNGQE 57

```





OY	816	EGRFSDSNLFNLSLPIGVKEFEKESQDNFSVDLSVPLINDKCKTALVIIS--GA	874
DJ	725	QGRFFESSNLITLSPGIGKFEPFNANDASTHVAALSPDIVANSNDPCTTSLLVS	928
OY	874	SWEYVANNLARQAQLQVRAGSHYAASFMEFVLGOEFVEWRSRRIRYNVDLGKGFQF	928
DJ	785	VWVTKANNNLARSAFMGLQAGNYLTSLSHNIETISQFGFELRGSSFRYYNDLGSKIQF	839
RESULT	11		
ID	086164	PRELIMINARY:	PRT: 928 AA.
AC	086164:		
DT	01-NOV-1998 (TrEMBLrel. 08, Created)		
DT	01-NOV-1998 (TrEMBLrel. 08, last sequence update)		
DT	01-MAR-2001 (TrEMBLrel. 16, last annotation update)		
DE	OUTER MEMBRANE PROTEIN 4 PRECURSOR.		
GN	OMP4 OR PMP.11.		
OS	Chlamydia pneumoniae (Chlamydophila pneumoniae).		
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.		
OX	NCB TaxID=83558;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CDC/CWL-029/VR-1310;		
RA	Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;		
RT	*Identification of two novel genes encoding outer membrane complex		
RL	associated surface layer proteins in Chlamydia pneumoniae.";		
RL	Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=VR1310;		
RA	Hierno K., Boesen T., Daugaard L., Knudsen K., Madsen A.;		
RA	Christiansen G., Birkelund S.;		
RT	"Chlamydia proteins containing the GGAI-repeat belong to a subfamily		
RT	of autotransporting pathogenicity factors.";		
RL	Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CWL029;		
RC	MEDLINE=9920606; PubMed=10192388;		
RA	Kalman S., Mitchell W., Marathe R., Iannet C., Fan J., Hyman R.W.;		
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.;		
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis";		
RL	Nat. Genet. 21:385-389(1999).		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN=J138;		
RC	MEDLINE=20330349; PubMed=10871362;		
RA	Shihai T., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.;		
RA	Shiba T., Ishii K., Hattori M., Kohara S., Nakazawa T.;		
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138		
RT	from Japan and CWL029 from USA.";		
RL	Nucleic Acids Res. 28:2311-2314(2000).		
DR	EMBL: AJ001311; CA04672.1;	-	
DR	EMBL: AJ133034; CAB37027.1;	-	
DR	EMBL: AB001628; AAD18599.1;	-	
DR	EMBL: AF002546; BA98658.1;	-	
DR	InterPro: IPRO03357;	-	
DR	InterPro: IPRO03368;	-	
DR	Pfam: PF02385; OMP; 1.		
DR	Pfam: PF02415; DUF145; 1.		
KW	Signal.		
FT	SIGNAL.	1	17
FT	CHAIN	18	928
FT	SEQUENCE	928 AA; 98903 MM; 788BCDD62C911402 CRC64;	POTENTIAL. OUTER MEMBRANE PROTEIN 4.

Query Match	37.2%	Score 1785;	DB 2;	Length 928;
Best Local Similarity	40.4%;	Pred. No. 6.1e-94;		
Matches 385;	Conservative 173;	Mismatches 344;	Indels 52;	Gaps 18;

Db	1	MKTSLIPWLVSVLA--PSCHQSLANELLSPDSSFNGNIDSGFPFPPKAT--TYSLT	56
Qy	61	GDITLQNGDSALTKGCFSDTTESLFAGKGYLSLFINIKSSAE--GAALSVTYTDKSL	115
Db	57	GVFVEFEPKGTPLDSCFCOTTDMLTFELGNHSLTFEGFDIAGTHAGAASTYANKNLTF	116
Qy	120	TFESSLITFLAASVYITPPSGKAVKCGGDLTFDNNGLILKQDYCEMGAISTKNLSL	179
Db	117	SGFSLSLSDSSPSYVTT--CGGTLSSAGVNLERILVAGNSTDADGAILGASFL	174
Qy	180	KNSTGISPEGKSSSATKGKGAICATGTVDTINTPAETFLSNSIAEAGAIINSTGCT	235
Db	175	TGTSGDALFNNSSST---KGALITTTAGARIANTTGYRFLSNIASTSGAIDDEGSI	233
Qy	240	ITGNTSLVFSSENSVATANGCAL-----SGDADVTISGNSVTFESGQAAVANGAIYAK	294
Db	232	LSNNKFLVPEGNAATT---GGALICNTAASGSPELIISNNKTLIFASVVAFTSGAIIAK	288
Qy	295	KTLTASGGGGGISFSNNIVGTTAANGAISITLAAECSLSAEAGDITFNGNAIYAT--TP	355
Db	289	KLTALS--GGFEFLRNWVSATP--KGALISIDASGELSLAEFGNITFVNTLTGGST	344
Qy	354	QTTKNSIDIGSTAKITNIRALISGSIFFPYPTIANTAADTDILNLNKADAGSTDYSG	413
Db	346	DIPKRNAINIGSGFTELRAKNITIFFPYDTISE--GTSSDLKINNGSAGALNPQG	403
Qy	414	SLVFGESKLSDEAVANLSTLKOPTVLNAGNIVLKRGTVLTKTGPTQASGVINDA	473
Db	404	TILFSGEITLADKLKAVADLKASSFPQPSLSGSKILLQKVTLESTSPQASGLIGMS	465
Qy	474	GTTLKASTEVTTLGSLIPVDSLGEKKVVIASAASKNALSGPILLDNOGNAYEND	533
Db	464	GTTLTSTAGSITITMIGINDVSLGKQVSLTAKGASKNIVSGKLNIDIEGNIYESHM	522
Qy	534	LKKTODESFVQLSA--LGIATITDVYA---VPPVATPTIKYQIGTGWMTVDDTSTPKTK	587
Db	524	FSHDQLFLSKLITVDADYDVNDISSLPVPAEDPNSEYEGQGGMNVMWTTDTAT--NTK	581
Qy	590	TATLTAMTNGVLPNPEROGPLVPNSLWMSFSFDIOAGVIRSLATLCSDBGFWAAGVAN	645
Db	582	EXTAAMWTGTGEVPSPEKRSALVCNLTLMGVFTDILSLQDLVEIGATGHEHKOGFWSSMTN	644
Qy	650	FLDKDKCKEKKRYKRSKGYAIGAQAQTCSENLISFACQLFGSDKDFLVAKNHTDYAG	705
Db	642	FLHKRGDEKRRKFRITSGGYVIGSSAHPKXDLFTFACHLFARDKOCFIANNHSSRTYGG	704
Qy	710	AFYIQTITEC-----SGFICGLDLKLPGSWNRPLVLEGOLAYSHVNDLTKRYT	755
Db	702	TLEFFHSHITLQPNYLRIGRAKFESESAIEKPP---REIPLALDVQVSPSHSDNMEHYTT	758
Qy	760	APPEVKGSMGNNAFMMMLGAS-----SHSYEYLHCPTFYAPYIKLNLTYIROSFEKG	814
Db	759	SLPEBEGSMNCCIAAGGIGLDLPVLSNHP---LFTFTLPQKVEVYVYSONSFESS	814
Qy	815	TGSRFSFDSNLFNLSLPIGVFEKSDCNDESYDLTSLVYDLIRNDPKCTTALVIGAS	874
Db	815	SGRGRFSIGRLNLSTIPGAFVYQODIDISTYDLSGFFVSDVYRNNPQSTATLVMPDS	874
Qy	875	WETVANNIARQALQVRAGSHAFSPMEFVLQGFVEVGRSSRIYVNDLGKQFQ	928
Db	875	WIRIGNGNISRQAFLLRGSNNYVYNSNCLEFHYAMELRIGSSRYNVDVGTLEF	928
RESULT	12		
Q9K299		PRELIMINARY; PRT; 949 AA.	
AC	Q9K299;		
DT	01-OCT-2000 (Tremblrel. 15, Created)		
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)		
DT	01-MAR-2001 (Tremblrel. 16, Last annotation update)		
DE	POLYMORPHIC MEMBRANE PROTEIN G FAMILY.		
GN	CP0302.		

OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,  
 RA Barry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AEO02192; AAF38159.1; -  
 DR TIGR: CP0302; -  
 DR InterPro: IPR003357; -  
 DR InterPro: IPR003368; -  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.  
 SQ SEQUENCE 949 AA; 101357 MW; A00B09E16C6998E3 CRC64;

Query Match 37.2%; Score 1785; DB 2; Length 949;  
 Best Local Similarity 40.4%; Pred. No. 6.3e-94;  
 Matches 385; Conservative 173; Mismatches 344; Indels 52; Gaps 18;

QY 1 MKSFSWLVLSSTLACTSCSTVFAATAENIGPSDFSGINTGTYTPKNTTGTIDYLT 60  
 DB 22 MKTSIPWLVLSVLA--FSCHLQSLANEELLSPDSENGMNDSTCTPTPKTSAT--TYSLT 77  
 QY 61 GDTLQNLGDSAAITKGFSDTETSLFAGKGYSLFLNLIKSSAE--GAALSVTTDKMLSL 119  
 DB 78 GDVFFYEPKGTPLSDSCFQOTDNLFLNGHSLTFCGIDAGHAGAASSTANKLTF 137  
 QY 120 TGTSSLTFLAAPSVTTPSGKAVKCGGLTDPDNGTTLFKODYCEENGGAISTKMLSL 179  
 DB 138 SGEFLSFDSPSTVTVT--GQGLTSSAGGVLENIRKLVYAGNFSYADGALIGASFL 195  
 QY 180 KNSGTSISFEGNKSATGKGAICATGVDITNNTAPTLEFNNIAAGAIINSTGCT 239  
 DB 196 TGTSGDALFSSNNSST--KGAIAATTAGARIANNITGVRLSIYSTSGCAIDDEGTST 252  
 QY 240 ITGNTSLVFSENSVTATAGNGAL----SGDADVTISGNQSVFSGNQAVANGAIYAK 294  
 DB 253 LSNKKFLYFEGNAKMT--GCAICNFKAGSGSPELIISNKKTLFASNVAVETSGAIHAK 309  
 QY 295 KLTLASGGGGISFSNNIVOGTTAGNGAISIILAGFCSLSAERGDITENGNAIVAT-TP 353  
 DB 310 KLAIS--GGFTEFLRNNVSSATP-KGAISIDASGELISLAETGNITFYRNTLTGTST 366  
 QY 354 QTKRNSIDIGSTAKTNLALISGHSLEFYDPTITANAASTDIDLNKKADAGSTIYSG 413  
 DB 367 DTPKRNAININGSNKFTLRNNAKNTTFEYDPTIS--GTSSDLKLTNNGSAGLNDYOG 424  
 QY 414 SIYFSGEKLSEDEKAVNDITSLKQPTVTLAGNLVLRKGTLDTKFOTAGASSYVMDA 473  
 DB 425 TIFESGTLADELKVADNLSKSTQPVSLISGKLLQKVTLESTFSGEAGSLGMDS 484  
 QY 474 GTTLKASTEEVTLTGSLIPVDSLGEGRKVVIAASAKNVALSGPILLDNOGNAYENHD 533  
 DB 485 GTLSTAGSITTLNGLINDISGLKQPVSLTAKGASNKYISGKLLNLDIEGNIYESHM 544  
 QY 534 LGTTOEFPVQAL-IGTATTTPVA---VPYATPHYGYOGTGWGTVDDDTASTKTK 589  
 DB 545 FSHDQFSLTKITIVADVDNVDISLIPVADPNSEYFQGMNWNMTDTAT--NTK 602  
 QY 590 TATLANTNTGYLNPBEROGLVFNLSLMSFSDIOALQGVERSALTLCSRGFVAGAVAN 649  
 DB 603 EATATYTKGTFVSPERKSLVYCNLTLMGYFTDRSLQQLVEIGATGEMHOGFVWSSMTN 662  
 QY 650 FLDDKKGERKRYRHSNGYVIGAAOTCSENLISFAFQGLFSGDKDLVAKNHTDTYAG 709

DB 663 FLHKTGDNENKGFHRHTSGGYVIGSAAHPKDDLTFFAFCHLFARDKCFIAHNNSRTYGC 722  
 QY 710 AFYIQTITEC-----SGFIGCLDLPGSMWKKPLVLEGQALAYSVNDIKTKYT 759  
 DB 723 TLFKHSHTLQPONYLRLGRKKFSESATKEP--KEIPLALDVQVSFSHDNRMEHTYT 779  
 QY 760 AYPEKSGWGNNAFNMWIGAS-----SHSYPEYLHCFDYAPYIKMLTYIRDSFSEK 814  
 DB 780 SLPESEGSWNECAGIGIDLPVLSNHP-----LFKFTLPQMKVEMYVSONSFESS 835  
 QY 815 TEGRSFDDSLFNLSPDIGVKEFEKSDCNDPSYDLISYVPLIRNDKCTTALVISGAS 874  
 DB 836 SDGRGFGISGLRLNLSIPVGAFFVOGDIGDSYTYDLGSEFVSDVYRNPNQSTAFALVMSPPS 895  
 QY 875 WETYANLQALQVBRAGSHYAFSPMEFVIGQFVEYRGSSRIYNDLGKFPQ 928  
 DB 896 WKIRGNLSRQAFLLRGSNNYVINSNCELGHYAMELRGSSRNINVDYGLKRP 949

## RESULT 13

P71132 PRELIMINARY; PRT; 847 AA.

AC P71132;  
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE POMp1A  
 OS Chlamydia psittaci (Chlamydia psittaci).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NCBI\_TaxID=83554;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;  
 RX MEDLINE=96406378; PubMed=8810511;  
 RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;  
 RT "Identification of a multigene family coding for the 90 kDa proteins  
 RT of the ovine abortion subtype of Chlamydia psittaci.";  
 RL FEMS Microbiol. Lett. 142:277-281(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;  
 RX MEDLINE=98187897; PubMed=9529048;  
 RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;  
 RT "Molecular cloning and characterization of the genes coding for the  
 RT highly immunogenic cluster of 90-kilodalton envelope proteins from the  
 RT Chlamydia psittaci subtype that causes abortion in sheep.";  
 RL Infect. Immun. 66:1317-1324(1998).  
 DR EMBL: U05942; AAC15921.1; -  
 DR InterPro: IPR003357; -  
 DR InterPro: IPR003368; -  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.  
 SQ SEQUENCE 847 AA; 90694 MW; 754C958E7F1179E CRC64;

Query Match 36.4%; Score 1745.5; DB 2; Length 847;  
 Best Local Similarity 41.9%; Pred. No. 9.7e-92;  
 Matches 402; Conservative 139; Mismatches 274; Indels 145; Gaps 27;

QY 1 MKSFSWLVLSSTLACTSCSTVFAATAENIGPSDFSGINTGTYTPKNT--TTGIDYT 58  
 DB 1 MKHPVWFLSSSLVLSNLYSE-EPDQKTLTSAHSYNGNTNSEPFPLSTSNNGITYT 59  
 QY 59 LRGDITLQNG-DSAAITKCFSDTETSLFAGKGYSLSLNLIK-SAEAAALSYR-TDK 115  
 DB 60 CTGNICLAYAGLDGSGSSCFDTAGNLSFLGNGYTLCDNTTQSSHPGAISVSGTNK 119  
 QY 116 NLSUTGSSLTFLAAPSVTTPSGKAVKCGGLTFDNDNGTTLFKODYCEENGGAISTK 175  
 DB 120 TLDISGS-----LFGCAVCPG----- 137  
 QY 176 NLSLKNSTGISFEGNKSATGKGAICATGVDITNNTAPTLEFNNIAEAGAIINST 235

```

Db 138 -----ATGY-----GAIKAV 147
QY 236 GNCITGNTSLVSENSVTATAGNGAL-----SGDADVTISGNQSVTFSGNOAVANGA 290
RT 148 GNTITKDNSSLVFHKNCST---GEGGAIQCKASSEAELEKLENNQNLVFAENSSSSGGA 204
QY 291 IYAKKLILASGGGGGIFSNNIYOGTTAGNGAISIL-AAGECGLSABAGITNGNAY 349
Db 205 IYAKKLIVS--GGPTLFSSNNVSASSP-KGCAICIDSGECSLITADLGITFDGKXII 261
QY 350 AAT--POTTKNSIDISTAKITNLRAISGHSIFEPDPTANTAADSDTLNKAAGN 407
Db 262 KTNQSGPTVTRNSIDLSSGKFTKLNAKEGFGIFFDITGG---GSDENIKKOD--- 314
QY 408 STDYSGSIVSEGEKLSDEDAKADNLSTLKPQVTLTAGNLVLRGVLTDTKGFOTAGS 467
Db 315 TVDYTGKIVFSGERLSDDEKKAANLKSDFKQPLKIGSGSLILDKGVTLLETSTQTEGA 374
QY 468 SVINDAGTTLKAST---EETVLTGLSTIPYDSLHG---KRVVIAASAASKNVALSGPIL 520
Db 375 TVVMDLGTTLQTPSSGGETTITLMDINAVSLGGGVAADPAKAEATESTVTIINA-VN 433
QY 521 LLDNQGNAVENHDLGKTODESFVOL-SALGATATTDVPAVPTVATPHYGYOGTGMATV 579
Db 434 LYDNGNANAYEPILAAQOPFALIEVRSGSSITKPTTNLENTYPPHHYOGQWMTYV- 492
QY 580 DDTASTPKTKATLMTNTGYLPNPEROGPLVPSNLGSPESDIOAGVIERALTLCS 639
Db 493 -KQSSAOKETATLTWEQGTGSPNPEROGSLPNTLWGSFSDIRAIQNLMDISVNGADYH 551
QY 640 RGFMAAGVANTLDKDKGKKRHRKSGGYAIGGAOTCSNLSLSPACOLFGSDKDFLY 699
Db 552 RGFVWVGSGNLFHSGSDTKRFRHNSAGYALGYAQTSPSDVESAPFQLEFGDKDYLV 611
QY 700 AKNMTDYAGAFYIOHTTECSGFIGCLLDKLPGSMHKLPLLEGOLAYSHVSNLTKRY- 758
Db 612 SKNSTYAGSIYOHISYMTW--NTLLQNTLG--AEPPLVLAQOLVCHASNNKNTMT 668
QY 759 -----TAYPEVKGWGNNAFNMLGASSHSYPEYLHCEPTVAPYKTLNLTITRODSFS 811
Db 669 DTVAPPKTYTESEIKDWCNDCFGVEFGAKAPETASL-LFDMYSPFVKQLQVHAHQDFK 727
QY 812 EKGH-EGRSPDLSLPLSLPIGYKFEKFSQCNDSYDLTISYVPLDRNDPKCTTALVI 870
Db 728 ENNSDOGRYFERNMLTJMSPIGVKLEKFSHKDTASYNLTAYADIVRSNPDCTASLLV 787
QY 871 S-GASMETYANLAROALQVRAQSHYAFSPMEFVLGOFEVEVRGSSRYVNDLGRKOF 928
Db 788 SPISAAMVYTKANNAIRHAFILQAGNYALALRNTELFSGFELNGSCRTYINIDLSKIOF 847

RESULT 14
P71133
ID P71133 PRELIMINARY; PRT; 846 AA.
AC P71133;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE POM91B PRECURSOR.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RX MEDLINE=98187897; PubMed=9529048;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RT "Molecular cloning and characterization of the genes coding for the
RT highly immunogenic cluster of 90-kilodalton envelope proteins from the
RT Chlamydia psittaci subtype that causes abortion in sheep.";
RL Infect. Immun. 66:1317-1324(1998).
DR EMBL; U65943; AAC15923.1; -.
DR InterPro; IPR003357; -.
DR InterPro; IPR003368; -.
DR Pfam; PF02385; OMP; 1.
DR Pfam; PF02415; DUF145; 1.
KW signal.
FT SIGNAL.
FT CHAIN 17 846 POTENTIAL.
FT POMP91B.
SQ SEQUENCE 846 AA; 90834 MW; 4CD31DC03C2964E CRC64;

```

```

RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RX MEDLINE=98187897; PubMed=9529048;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RT "Molecular cloning and characterization of the genes coding for the
RT highly immunogenic cluster of 90-kilodalton envelope proteins from the
RT Chlamydia psittaci subtype that causes abortion in sheep.";
RL Infect. Immun. 66:1317-1324(1998).
DR EMBL; U65943; AAC15923.1; -.
DR InterPro; IPR003357; -.
DR InterPro; IPR003368; -.
DR Pfam; PF02385; OMP; 1.
DR Pfam; PF02415; DUF145; 1.
KW signal.
FT SIGNAL.
FT CHAIN 17 846 POTENTIAL.
FT POMP91B.
SQ SEQUENCE 846 AA; 90834 MW; 4CD31DC03C2964E CRC64;

Query Match 35.5%; Score 1701; DB 2; Length 846;
Best Local Similarity 41.2%; Pred. No. 3,4e-89;
Matches 396; Conservative 133; Mismatches 263; Indels 150; Gaps 26;

QY 1 MKSQFNLVLSSTLACTSCSTVFA-ATAENIGPSDSFGSTNTGYTPKNTTGTIDYTL 59
Db 1 MKHPVYFLLSSLSLASNSLS-FAQVTNETLSSDSYNGNVTDEFEVETTSAGIATC 58
QY 60 TGDITLONLDSALITGCSDDTTESSLPAGKYSLSFLNKS-SABGALSVT-TDKNL 117
Db 59 EGNVCISYAGDPSLKNKSCSEETENLFTGNGYTLCEFDNITQSSHPGALSAGVTNLT 118
QY 118 SLNGESSLFLAAPSVTTPPSGKAVKCGDGLFDNNGTTLFPODCEENGALISRNKL 177
Db 119 DISGFS-----LSCAVC----- 131
QY 178 SLKNSSTSISEPNKSSATGKGAICATGVDITNTNAPTLLFSNNIAEAGAINSTGN 237
Db 132 -----CPPTGTG-----GAIQKGT 147
QY 238 CTITGNTSLVSENSVTATAGNGAL-----SGDADVTISGNQSVTFSGNOAVANGAIY 292
Db 148 TTLKDNSSLVFNKNCSTA---EGGAIQCKSSSSYAELEKLENNKMLVSENSKKEGGAIIY 204
QY 293 AKKTLTASGGGGGIFSNNIYOGTTAGNGAISIL-AAGECSLSAEGDITFNENATYAT 351
Db 205 ADKLITVS--GGPTLFSSNNVSASHSSPRGAIKIKDSDGSLVLANGLDITFDKNTIT 262
QY 352 T--POTTKNSIDIGSTAKITNLRAISGHSIFEPDPTANTAADSTPLNKNKADGNS 409
Db 263 NGSPPTVTRNSIDLSSGKFTKLNAKEGFGIFFDPTI-ANTGSGTELELNKTESD-----T 317
QY 410 DYSGSIVSEGEKLSDEDAKADNLSTLKPQVTLTAGNLVLRGVLTDTKGFOTAGSSV 469
Db 318 TYTGKIVFSGERLSDDEKKAANLKSDFKQPLKIGASLVKQDVTLEAKKITQTKGSTV 377
QY 470 IMDAGTTLK---ASTEVTLTGLSIPYDSLHG---KRVVIAASAASKNVALSGPILLD 523
Db 378 VMDIGTTLQTPSSGGETTITLMDINAVSLGGGVAADPAKATLTAQALISIA-VNLVN 436
QY 524 NQGNAYVENHDLGKTODESFVOLSAIGATTTDVPVAV-----TVATPPTHYGYOGTGMAT 577
Db 437 TDSNTYEDPLISAKSFSAT-----TATSSSTVTPPEPNTLKNYTPPHYGYOGNMTVT 490
QY 578 VVDDTASTPKTKATLMTNTGYLPNPEROGPLVPSNLGSPESDIOAGVIERALTLIC 637
Db 491 W--KQSSAOKETATLTWEQGTGSPNPEROGSLPNTLWGSFSDIRAIQNLMDISVNGAD 548
QY 638 SDRGFMAAGVANTLDKDKGKKRHRKSGGYAIGGAOTCSNLSLSPACOLFGSKDF 697
Db 549 YSRNFWFVSSLANFLNKGSDTKRFRHNSAGYALGYAQTSPSDVESAPFQLEFGDKDY 608
QY 698 LVAKNHTDYAGAFYIOHTTECSGFIGCLLDKLPGSMHKLPLLEGOLAYSHVSNLTKR 757
Db 609 FVSKNSSTIYAGSIYOHISYMTW--NTLLQNTLG--AEPPLVLAQOLVCHASNNKNTMT 665

```

QY 758 YT-----AYPEKSGMGNNAFNMMLGASSHSYPEYLHCFDPTAYIKLNTIYRDS 809  
 DB 666 MTNTYPTKNTYPSIKDKWGDGCGVEFGAKAPLETASTL-LFDMYSPFVLIQVHAQDD 724  
 QY 810 FSEKGT-EGRSFDDSNLFNLSPILGVKFEKFSQDNDSYDLTSLSYVDLIRNDPKCTTAL 868  
 DB 725 FKENSNDQGRYFESNLTJNLMSPIGVKLEKFSHKDTASYNLTLAYADIVASNDPCTASTL 784  
 QY 869 VIS--GASMETYANNLAROALQVRAAGSHYAFSPMEVLQGFVEFVGRSSRIYNDLGKF 926  
 DB 785 LVSTSAVWVTKANNLRAHAFITLQAGVTLALTRNTELSQFPELRSGCRITNIDLSKI 844  
 QY 927 QF 928  
 DB 845 QF 846

RESULT 15  
 Q923A1 PRELIMINARY: PRT: 841 AA.  
 AC Q923A1:  
 DT 01-MAY-1999 (TEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)  
 DE OUTER MEMBRANE PROTEIN 7 (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).  
 GN OMP7 OR PMP.2 OR CP0761.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_taxid=8358;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kaimen S., Mitchell W., Marathe R., Lammel C., Pan J., Hyman R.W.,  
 RA Olinier L., Grimwood J., Davis R.W., Stephens R.S.,  
 RT Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VL1310;  
 RA Hjertoe K., Boesen T., Daugaard L., Knudsen K., Madsen A.,  
 RA Christensen G., Birkelund S.;  
 RT "Chlamydia proteins containing the GAT-repeat belong to a subfamily  
 of autoexporting pathogenicity factors.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwynn M., Nelson W., DeBoy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genomic sequences of Chlamydia trachomatis MoPn and Chlamydia  
 pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AEO01586; AAD18172.1; -;  
 DR EMBL: AUI33035; CAB37083.1; -;  
 DR EMBL: AEO02235; AAF38561.1; -;  
 DR TIGR: CP0761; -;  
 DR InterPro: IPR003357; -;  
 DR InterPro: IPR003368; -;  
 DR Pfam: PF02385; OMP.1;  
 DR Pfam: PF02415; DUF145; 1.  
 SQ SEQUENCE 841 AA; 89600 MW; 906460D0678D24C CRC64;

Query Match 29.3%; Score 1406.5; DB 2: Length 841;  
 Best Local Similarity 34.3%; Pred. No. 2.1e-72;  
 Matches 330; Conservative 162; Mismatches 314; Indels 157; Gaps 23;

QY 1 MKSOFWLVLSLTLACPTSCSTVFAATAENIGPSDFSGSTNTGYTPK--NTTGYIDYT 58  
 DB 1 MKTIPRLRL-SLVPTLSMNLGAATEELLSANSPDGTSTTSFSKTSSTADTINY 59  
 QY 59 LTGDITLQNL---GDSALLKGCFSOTTES--LSFAGKISLSFLNI-KSASGAL-SV 111  
 DB 60 FKDSVYLENVPKGTGEQSTG--CFKNDMAAGDILNFGGGSFPFSNIDATAGGAAGSE 117  
 QY 112 TTDKSLTGFSSLTFLAAPSVTTPPSGKAVKCGDILFDNNGTILFKDYCEENGGA 171  
 DB 118 AANKVTLSGFSALSFKSPASTYT-----NGL----- 145  
 QY 172 ISPTKSLSKNSTGSISEPNKSSATGKGAICATGTVDITNTAPTTLFSNNIAEAAGA 231  
 DB 146 -----GA 147  
 QY 232 INSTGCTITGNTSLVSEMSVTATANGGALSADADVTISGNSVTFSGNOAVANGAI 291  
 DB 148 INVGKSLTLDNKKVLLIDNFST--GDGAINCAGSLKTAANKSLSEIGNSSSTRGAI 204  
 QY 292 YAKKLTLAGGGGGISPSNNIVOG---TTAGNGAISILIAEGCSLSAEGDITFNGNA 347  
 DB 205 HTKNLTLSGG-----ETLFGONTAPTAAGKAGALAIADSGTSLSDSGDILFPGNT 257  
 QY 348 IVATTPQTRKNSIDIGSTAKITNLRAISGHSIFEPDPTANTPAADSTDTLNLKADAGN 407  
 DB 258 IGAT--GTVSHSAIDIGTSKAITALRAAQCHTIFYDPIVTGSTVADALININSPDTC 315  
 QY 408 STDYSGSIVSGEKLSDDEKAVADNLTSLKOPYTLTAGNLVLRGVTLLTKGFTQAGS 467  
 DB 316 NKEYGTIVFSGEKLTEAEKADKERNRTSKLLQNVAFKNGVTLVKGDVLSANGSDQANS 375  
 QY 468 SVINDGTTKASTEEVYTLGSLIPVDSLEGKVVYLAASAASKNVALSGPILLDNOGN 527  
 DB 376 KLINDGTSIVANTSEIETLNLEINIDSLNGKKIKSATQKDIRIDRPVYLAISDES 425  
 QY 528 AYENHDIGKTQDF-SEYQLSA-LGATATTTVPVAVPTATPHYGTGCTGCTGCTGCTGCT 585  
 DB 436 FYONGFLNEDHSYDGLLELAGKDIVISADSRSDAVQSP--YGYOGKWTINSTD---- 489  
 QY 586 PKTKATLANTNTGYLPNPRGGLVAVNSLNGSPSDIOATOGYIERSALTLCSDRGPMA 645  
 DB 490 --DKKATVSWAKOSFNPTAQDEAPLVNLMGSLIDRSQNTLELGTGECAPTEKRFWA 547  
 QY 646 GVANFLDKDKGKRRKRYHSGGAYALGAATCSENLISFAFQLEFGSDKFLVAKNHTD 705  
 DB 548 GISVTLHRSRGRENKRHRHSGGAVVAGASTRMFGDITLSGFQQLFARDQDYFMNTFAK 607  
 QY 706 TYAGAFYIOHTTECSGFIGCLDK-----LPGSMHAKPLYLEGOLAYSHVNDLTKY 758  
 DB 608 TYAGSLRLQHDASLYSVYSTILLEGGLREILLPVSKTLCSFYGQLSYGHTRHMTES 667  
 QY 759 TAYPEV-----KSGMGNNAFNMMLG-----ASSHSYPEYLHCFDPTAYIKLNTIYR 806  
 DB 668 LPPPPPLSLTDHRSWGSYVWAGELGTVAVENTSGKGF-----FOETTFVAVQAVYAR 721  
 QY 807 QDSSEKTEGRSFDSDSNLFNLSPILGVKFEK--FSDCNDSYDLTSLSYVDLIRNDPKCT 865  
 DB 722 QDSFVELGALSIPDSOHLVNLAIPLIGKLEKRA--EQYHVAVMYSPDVCRSNPKCT 778  
 QY 866 TALVYSGASMETYANNLAROALQVRAAGSHYAFSPMEVLQGFVEFVGRSSRIYNDLGK 925  
 DB 779 TTLISNGSWKTKGSLNAROAGIVQAGSFRSLAAALFLGNFGFEWGSRSRYNVADGSK 838  
 QY 926 QF 928  
 DB 839 IKF 841

Search completed: October 2, 2001, 03:31:02  
 Job time: 242 sec

BEST AVAILABLE COPY

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 00:43:56 ; Search time 18674.8 Seconds  
(without alignments)  
2527.870 Million cell updates/sec

Title: US-09-446-677B-5  
3052  
Sequence: 1 ATGGGATTTCGCTCTGCGG.....GTTCCTTGAGTCCTTGCA 3052

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 773874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_v1:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v11:\*  
59: gb\_v12:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_ro1:\*  
95: gb\_ro2:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	3052	100.0	3052	9 A81831	A81831 Sequence 5
2	3052	100.0	26920	2 CPN133035	AJ133035 Chlamydia
3	3047.2	99.8	10766	1 AE001585	AE001585 Chlamydia
4	3045.6	99.8	299650	2 AP002545	AP002545 Chlamydia
5	3044.4	99.8	10044	1 AE002237	AE002237 Chlamydia
6	1192	39.1	1200	9 A81853	A81853 Sequence 27
7	272.6	8.9	12127	1 AE002235	AE002235 Chlamydia
8	271	8.9	16448	1 AE001587	AE001587 Chlamydia



9	270	8.8	2787	9	A81835	A81835 Sequence 9
10	264.8	8.7	4926	2	CPU2499	U72499 Chlamydomph
11	260.8	8.5	10026	1	AE002193	AE002193 Chlamydomph
12	260.8	8.5	15068	1	AE001627	AE001627 Chlamydia
13	260.8	8.5	300650	2	AP002546	AP002546 Chlamydomph
14	252.6	8.3	3200	9	A81827	A81827 Sequence 1
15	252.6	8.3	6030	2	CPOMP54	AJ001311 Chlamydia
16	252.6	8.3	10757	1	AE001628	AE001628 Chlamydia
17	252.6	8.3	12676	1	AE002192	AE002192 Chlamydomph
18	252.6	8.3	17280	2	CPN133034	AJ133034 Chlamydomph
19	233.2	7.6	12676	1	AE002192	AE002192 Chlamydomph
20	233.2	7.6	300650	2	AP002546	AP002546 Chlamydomph
21	232.6	7.6	10757	2	AE001628	AE001628 Chlamydomph
22	231.6	7.6	2815	9	A81829	A81829 Sequence 3
23	231.6	7.6	17280	2	CPN133034	AJ133034 Chlamydia
24	226.6	7.4	2793	9	A81841	A81841 Sequence 15
25	222.8	7.3	3000	9	A81851	A81851 Sequence 25
26	222.8	7.3	6030	2	CPOMP54	AJ001311 Chlamydia
27	217.4	7.1	2757	9	A81837	A81837 Sequence 11
28	184.6	6.0	2787	9	A81839	A81839 Sequence 13
29	178	5.8	2838	9	A81849	A81849 Sequence 23
30	159.4	5.2	10574	1	AE001586	AE001586 Chlamydia
31	150.4	4.9	2526	9	A81833	A81833 Sequence 7
32	150	4.9	840	9	A81843	A81843 Sequence 17
33	138.6	4.5	6110	2	CPU65942	U65942 Chlamydomph
34	126.6	4.1	6234	2	CPU65943	U65943 Chlamydomph
35	106	3.5	10695	1	AE001629	AE001629 Chlamydia
36	106	3.5	15789	1	AE002191	AE002191 Chlamydomph
37	105	3.4	1830	9	A81855	A81855 Sequence 29
38	64.6	2.1	14177	1	AE001360	AE001360 Chlamydia
39	64	2.1	14482	1	AE002293	AE002293 Chlamydia
40	63.2	2.1	1454	9	A81845	A81845 Sequence 19
41	47.8	1.6	10001	1	AE002248	AE002248 Chlamydomph
42	47.8	1.6	11781	1	AE001676	AE001676 Chlamydia
43	47.8	1.6	32585	2	AP002548	AP002548 Chlamydomph
44	45	1.5	180354	61	AC009689	AC009689 Homo sapi
45	44.8	1.5	1141	10	AX083744	AX083744 Sequence

## ALIGNMENTS

```

RESULT 1
A81831 A81831 3052 bp DNA PAT 21-JAN-2000
LOCUS Sequence 5 from Patent WO9858953.
DEFINITION A81831
ACCESSION A81831
VERSION A81831.1 GI:6731866
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3052)
AUTHORS Madgen, A. and Birkelund, S.
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL Patent: WO 9858953-A 5 30-DEC-1998;
MADSEN ANNA SOPHIE (DK); BIRKELUND SVEND (DK)
FEATURES
Source 1..3052
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 875 a 603 c 653 g 921 t
ORIGIN

```

```

Query Match 100.0%; Score 3052; DB 9; Length 3052;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ATGGCAATTCGCTGCGGATTCCTAGTCTTCTTACATTCGCTGCTTC 60
DB 1 ATGGCAATTCGCTGCGGATTCCTAGTCTTCTTACATTCGCTGCTTC 60

```

QY	61	GACACCTCTTTGAGTCTACTACGATTTCTTTAAACCCAGAAAGTAGTTTCATGAGAT	120
DB	61	GACACCTCTTTGAGTCTACTACGATTTCTTTAAACCCAGAAAGTAGTTTCATGAGAT	120
QY	121	AGCAGAAATGCAAGCTTCTTAATGTTCAAGCTGGGAGTCTATAGCTTACTGGT	180
DB	121	AGCAGAAATGCAAGCTTCTTAATGTTCAAGCTGGGAGTCTATAGCTTACTGGT	180
QY	181	GATGTCATATCTAAGCTGATTAAGCTGATTAATTAAGCCGCTCATGTGACC	240
DB	181	GATGTCATATCTAAGCTGATTAAGCTGATTAATTAAGCCGCTCATGTGACC	240
QY	241	TCAGAAAGTGTGACGTTCCGAGAAATCATGCGTTATTTTAATATTTCTCTA	300
DB	241	TCAGAAAGTGTGACGTTCCGAGAAATCATGCGTTATTTTAATATTTCTCTA	300
QY	301	GGACATCAAGAAAGAGGGGTGCTACTTTGTGGCAAGTCCCTAACACGGACGTTT	360
DB	301	GGACATCAAGAAAGAGGGGTGCTACTTTGTGGCAAGTCCCTAACACGGACGTTT	360
QY	361	TCGCGGTTCCACGCTCTCTTTATTCAGAGCCCGGAGATTTAAAGACAGGATGT	420
DB	361	TCGCGGTTCCACGCTCTCTTTATTCAGAGCCCGGAGATTTAAAGACAGGATGT	420
QY	421	CTCTATTCAAAAATGCACTTATGCTTTAAACAATATATGATGCGTTTGAACAAAC	480
DB	421	CTCTATTCAAAAATGCACTTATGCTTTAAACAATATATGATGCGTTTGAACAAAC	480
QY	481	CAAAGTAAGCTTAAAGCGGAGCTATTAGTGGGCGCAATGTTACTATAGGCACTAC	540
DB	481	CAAAGTAAGCTTAAAGCGGAGCTATTAGTGGGCGCAATGTTACTATAGGCACTAC	540
QY	541	GATTCGCTCTCTTCTATGAGATGAGCAGCCACTTTTGGAGGTCATTCCTTCAAGT	600
DB	541	GATTCGCTCTCTTCTATGAGATGAGCAGCCACTTTTGGAGGTCATTCCTTCAAGT	600
QY	601	CCCTACAGATTCAGTAATATCAGCAGAGATATAGATTTGACAAATATCTGCCAAGAT	660
DB	601	CCCTACAGATTCAGTAATATCAGCAGAGATATAGATTTGACAAATATCTGCCAAGAT	660
QY	661	GGTTCGAGGGGCTTGTACTCCGATGGATGATGATGATGATGATGATGATGATGAT	720
DB	661	GGTTCGAGGGGCTTGTACTCCGATGGATGATGATGATGATGATGATGATGATGAT	720
QY	721	CTATTCGAGAAATAGGAGCTTACTGCTATAGTAAAGGAGGAGGCTGCTTGT	780
DB	721	CTATTCGAGAAATAGGAGCTTACTGCTATAGTAAAGGAGGAGGCTGCTTGT	780
QY	781	CTTCCACTTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	840
DB	781	CTTCCACTTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	840
QY	841	TTAGCTTTGAAGAAACCATTCATTAATGGGAGGAGGAGGAGGAGGAGGAGGAGGAG	900
DB	841	TTAGCTTTGAAGAAACCATTCATTAATGGGAGGAGGAGGAGGAGGAGGAGGAGGAG	900
QY	901	AGCATCTCTCAGAGAGCTCTACTATTTATCAATATATATATATATATATATATAT	960
DB	901	AGCATCTCTCAGAGAGCTCTACTATTTATCAATATATATATATATATATATATAT	960
QY	961	AATTTAGTGTGACCTATTCATTTAGTGTGACCTATTCATTTAGTGTGACCTATTC	1020
DB	961	AATTTAGTGTGACCTATTCATTTAGTGTGACCTATTCATTTAGTGTGACCTATTC	1020
QY	1021	GGACATTAATCAATTCAGAAACCGAGAGAGTCCGTTTGAATGGATCCATCTT	1080
DB	1021	GGACATTAATCAATTCAGAAACCGAGAGAGTCCGTTTGAATGGATCCATCTT	1080
QY	1081	TTTCAAAATGCTAAATTCCTGAATATACAGGAGAAATGATGCTCTATAGATTTTAT	1140
DB	1081	TTTCAAAATGCTAAATTCCTGAATATACAGGAGAAATGATGCTCTATAGATTTTAT	1140
QY	1141	GATCTATTATCTTCTGAGAGAGTGGCTACCCCAATTAATATACAGGAGATCTTAA	1200

Db 1141 GATCTATTACTTCTGAGAGCATGGCTCTACCCATTTGAATATCAAGGAGATCTTAA 1200  
 QY 1201 AATAAAGAGTACACAGGAGGCATACCTCTTTCTGAGAGAAAGAGTCTAGCAAAAGATCTT 1260  
 Db 1201 AATAAAGAGTACACAGGAGGCATACCTCTTTCTGAGAGAAAGAGTCTAGCAAAAGATCTT 1260  
 QY 1261 AGGATTTTAAATCTACAAATCCCTCAGAAAGTCAACCTGCTGCAGAGTACTTAAATTTT 1320  
 Db 1261 AGGATTTTAAATCTACAAATCCCTCAGAAAGTCAACCTGCTGCAGAGTACTTAAATTTT 1320  
 QY 1321 AAAGAGGGGGCCGAGTACAGTTCCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 1380  
 Db 1321 AAAGAGGGGGCCGAGTACAGTTCCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAG 1380  
 QY 1381 TTGATTTTAAAGACCAAACTGATAGCTCTTAAGAGAGACATTTGCCATCAGAGCTGCGC 1440  
 Db 1381 TTGATTTTAAAGACCAAACTGATAGCTCTTAAGAGAGACATTTGCCATCAGAGCTGCGC 1440  
 QY 1441 ATGATATATAGATAGCTTAAGCTCATCCCTCAGAGAGTCTGATTTAAAGCAACCGCA 1500  
 Db 1441 ATGATATATAGATAGCTTAAGCTCATCCCTCAGAGAGTCTGATTTAAAGCAACCGCA 1500  
 QY 1501 AATAAAGAGATATCCGTCAGAGGACTCTATAGACTTATCTGCGCTAGCGCAATGCTTAT 1560  
 Db 1501 AATAAAGAGATATCCGTCAGAGGACTCTATAGACTTATCTGCGCTAGCGCAATGCTTAT 1560  
 QY 1561 GAAGATCTCAGAAATGAGAAATTCACAGAGCTTCCCTGCTCTCTTTAGAGCCCTGAGCC 1620  
 Db 1561 GAAGATCTCAGAAATGAGAAATTCACAGAGCTTCCCTGCTCTCTTTAGAGCCCTGAGCC 1620  
 QY 1621 GGGGCTAGTGTGACTGATCTGATCTGAGAGATTTCTACCGGATTAAGTCCCATTAATGCTTT 1680  
 Db 1621 GGGGCTAGTGTGACTGATCTGATCTGAGAGATTTCTACCGGATTAAGTCCCATTAATGCTTT 1680  
 QY 1681 CAAGGCAATTTGGAATTTAGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
 Db 1681 CAAGGCAATTTGGAATTTAGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
 QY 1741 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800  
 Db 1741 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800  
 QY 1801 TGGGGGAATGCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860  
 Db 1801 TGGGGGAATGCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860  
 QY 1861 TTACAGACAGATCGAGGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920  
 Db 1861 TTACAGACAGATCGAGGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920  
 QY 1921 TCCGAGAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1980  
 Db 1921 TCCGAGAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1980  
 QY 1981 GAGATCAACCTAGCACTATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040  
 Db 1981 GAGATCAACCTAGCACTATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040  
 QY 2041 GACTATGCGGTTTCCCAACAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2100  
 Db 2041 GACTATGCGGTTTCCCAACAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2100  
 QY 2101 ACAACCTCCCTAGGGAATATTTTCCGTTATGCTTCCGTTATGCTTCCGTTATGCTTCCG 2160  
 Db 2101 ACAACCTCCCTAGGGAATATTTTCCGTTATGCTTCCGTTATGCTTCCGTTATGCTTCCG 2160  
 QY 2161 ATTCTCTCAAAAGGTTTCTCAAAATCCCTTATGATTTTCTATTTTCTGCTTAT 2220  
 Db 2161 ATTCTCTCAAAAGGTTTCTCAAAATCCCTTATGATTTTCTATTTTCTGCTTAT 2220  
 QY 2221 GGTATGCGCAACCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2280  
 Db 2221 GGTATGCGCAACCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2280

Db 2221 GGTATGCGCAACCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2280  
 QY 2281 TGGAGAAACAAATTTGTTGGCTATATAGAGTGGGAGGAGCATGCTTATTTGATTTGAG 2340  
 Db 2281 TGGAGAAACAAATTTGTTGGCTATATAGAGTGGGAGGAGCATGCTTATTTGATTTGAG 2340  
 QY 2341 AACGAGACTTTTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2400  
 Db 2341 AACGAGACTTTTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2400  
 QY 2401 CAGGAGACTTTTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2460  
 Db 2401 CAGGAGACTTTTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2460  
 QY 2461 TCGATTTCTGATCTCTAGGAGATAGGCTTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2520  
 Db 2461 TCGATTTCTGATCTCTAGGAGATAGGCTTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2520  
 QY 2521 TATGACTTTAGTTTCTCTATATCTGATATTTTCCGTTAGAGATCCCTCATGGAAGCT 2580  
 Db 2521 TATGACTTTAGTTTCTCTATATCTGATATTTTCCGTTAGAGATCCCTCATGGAAGCT 2580  
 QY 2581 GCTCTGATATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2640  
 Db 2581 GCTCTGATATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2640  
 QY 2641 TTTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2700  
 Db 2641 TTTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2700  
 QY 2701 GGAAGTATAGAGTCCGCGCCCATGCTAGAGATTAATTAATTAATTAATTAATTAATTAAT 2760  
 Db 2701 GGAAGTATAGAGTCCGCGCCCATGCTAGAGATTAATTAATTAATTAATTAATTAATTAAT 2760  
 QY 2761 CGTTTGAAGAGTTTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820  
 Db 2761 CGTTTGAAGAGTTTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820  
 QY 2821 ATGATATATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2880  
 Db 2821 ATGATATATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2880  
 QY 2881 TTAAGGATATCTGATATCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2940  
 Db 2881 TTAAGGATATCTGATATCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2940  
 QY 2941 CTAACCTTCTGATTTTATCGAATCCCTTACGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 3000  
 Db 2941 CTAACCTTCTGATTTTATCGAATCCCTTACGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 3000  
 QY 3001 TCTCTAACGAATCATAGGAGGATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3052  
 Db 3001 TCTCTAACGAATCATAGGAGGATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3052

RESULT 2  
 CPN133035 26920 bp DNA BCT 27-FEB-2001  
 LOCUS  
 DEFINITION Chlamydia pneumoniae GA1-protein family cluster B, genes omp6,  
 hb1, hb2, hb3, hb4, hb5, hb6, hb7, omp7, pseudogenes omp8, omp9,  
 omp15, strain VR1310.  
 ACCESSION AJ133035  
 VERSION AJ133035.1 GI:4455890  
 KEYWORDS HB1 gene; HB2 gene; hb3 gene; hb4 gene; hb5 gene; hb6 gene; hb7  
 gene; omp15 gene; omp6 gene; omp7 gene; omp8 gene; omp9 gene; outer  
 membrane protein 6; outer membrane protein 7; pseudogene.  
 SOURCE Chlamydia pneumoniae.  
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 REFERENCE 1 (bases 1 to 26920)  
 Daugaard, L., Hjerno, K., Knudsen, K., Madsen, A.S., Christensen, G.  
 and Birkebaard, S.  
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 26920)  
AUTHORS Boesen, T.  
TITLE Direct Submission  
JOURNAL Submitted (21-JAN-1999) Boesen T., Department of Medical Microbiology and Immunology, University of Aarhus, The Bartholin Building, DK-8000 Aarhus, DENMARK

FEATURES  
source Location/Qualifiers  
1. .26920  
/organism="Chlamydia pneumoniae"  
/strain="VR1310"  
/db\_xref="taxon:83558"  
498. .3266  
/gene="omp6"  
498. .575  
/gene="omp6"  
498. .3266  
/gene="omp6"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein 6"  
/protein\_id="CAB37075.1"  
/db\_xref="GI:4455891"  
/db\_xref="SPTREMBL:Q924H9"  
/translation="MKFSLGFPPLVPSLTLSVPTSLGATTSISLPEDSEPHGDSONA  
EESYVQADGVSILTGDSISVNDNSALKACFNVTSGSVTFAGNHGILFNNISGT  
TRGAVLCCDDPOATAREFSGFTLSFIOSPGDIKEGCLYKSNALMLNNYVVR  
OSKTKGAGISANVTIGVNDVSFYONATFGAIIHSSGPLOIAVNOAIEFAONTA  
KSGSGALSDSDIDIDONAAVILFRPNEALTTAICGAGVCLPTSGSPVPIYTS  
DKOIVPEBNHNGSGALYAKLSISGGPTLFNNISYANSQNGAIAIDTGEI  
SLSAEGTTTFQGNRTSLPFLNGHILQNAKFLQARNGCSIEFDLPITSDASTO  
LNIINDPKKEYGTTLFSGEKSILANDPDEFTSLPQNVNLSAGLYIKEGAEVYSK  
FTQSPGSHLVLDIGTKLASKEDIAITGLAIDISLSSSTAIVIKANANKOISVTD  
STELISPTGNATEDLRMNSOTFPLISLEPGAGAGSVYTAGDFLPVSPHYGQWKL  
AWTGNGKNGEFGFMKINPKPRKGNLVNINLVNNAVNSLMQOETHSSILQTD  
RGIMDIGNPFHVASSEDNITRRHNSGYYIVNNETPRKTYTSKABQSLSRKDY  
AVSNNEPRYLGSTLYQYTTSLGNTFRYASRNPVNGILSRFLQNPIMATHFLCAY  
GATNDKMTDVANFPVKNNSMKNWMAIECGSMPLVFNGLRFGALPEKKLDLY  
AGVDPKETTADGRRFNSGLTSISVPLGIREKLAISODVLYDFSFSYIPIDFKDP  
SCFALVLSGDSMLVPAHVSHRAVVGSGTGRHFNDDYTELLRGSIIECRPHARNVI  
NCGSKFRP"

gene  
3976. 6870  
/gene="hb1"  
3976. .6870  
/gene="hb1"  
/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB1 protein"  
/protein\_id="CAB37076.1"  
/db\_xref="GI:4455892"  
/db\_xref="SPTREMBL:Q923C8"  
/translation="MRKRLRIALIVLIALSIIILAGVVLTVAIPLGLSSVTSPPAGMG  
ACALGCMALGIDVLKKREPIYLASVTTPTGSGPRSGISISGADSTISLPTLY  
LDEGHPQSKRLRIALIVLIVFSITLISGLVLLVAIPGLSSVTSPPAGMACALGC  
VMAIGIDVLKKREPIYLASVTTPTGSGPRSGISISGADSTIRSLPTPLDGHG  
QSMKRLRIALIVLIVFSITLISGLVLLVAIPGLSSVTSPPAGMACALGCALG  
IDVLKKREPIYLASVTTPTGSGPRSGISISGADSTIRSLPTPLDGHG  
VESHEENKSLPYDGHGEKTKHOIRVYRSLSKAMPFLDIRLFPFEEFFELSTAR  
KLIDLATTVIERKTLTLEOLERNNRKAPSVLYQOSTFKTIIDNEKLAAMKMLISKS  
ICRPTITFNEHGVAKSLHKNVALLKVRISLQKSTRDIMGKSAAMKILHGPFF  
SLEDKQKTIKMEHAEMESLSSYKRVFLALSDENVDDPPSDKWDLSGIPQDLASE  
ISREOMQKKAHLKHQESLITQARDLTDQSKENQKELKAQDEYISWMEKGVKFI  
ERVOERAIQIKLYENILREBEETGOEVTPTVOGTTASSDLTDILGIESAKEDFI  
ONESCVCVLRSHHEVMSWEKOEYGPKKKEKRODMGSLERFPHIIEELVQSDYS  
KHLSTFKVNNKKEVOYAKFRUKYLESDEGLAOTESAESLITDEEPLITATKALE  
KAVRGSLLCASKAKPFIIEEDPFOQSDQLRLTLRLQPAKASLEEKIKRFSNLE  
NDIAERLLKLESKQTFERRAGIAGVIREIAVESTYDLRSITNTWECTPSEKRYEYML  
NYVNEKRRARTRYEVTQRYRDFKMALEAMQFNEALLOEELSIQADSE"

gene  
7150. .8058  
/gene="hb2"  
7150. .8058  
/gene="hb2"

/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB2 protein"  
/protein\_id="CAB37077.1"  
/db\_xref="GI:4455893"  
/db\_xref="SPTREMBL:Q925Q3"  
/translation="MOVFLSPOLPPPPRSLGVSCSKLRSISITLLVLGLLTGIP  
GLTAGISGAGISALGVLVIGLLFLVREVPYVSEELPRGVSVTPSEEPALF  
KAKPEPMKTIIDRLPKELDQIDTIOFVFAFLPKDKRYEPLAPAKRLVAD  
VYKDMSEFLDIQRLNEEATYVEHCODPLNNIAYETFSOELRDYTAGCGTUPS  
GDARADRLKRSYKVEWDRFMTWKSWEASVMLDSYGVARELLFKAVGLSESYKI  
LFKSYRADYFEECEKAKIORDGRKWL"  
8185. .9492  
/gene="hb3"  
8185. .9492  
/gene="hb3"  
/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB3 protein"  
/protein\_id="CAB37078.1"  
/db\_xref="GI:4455894"  
/db\_xref="SPTREMBL:Q923C0"  
/translation="MMLDRYADKFLIREKEERMERHELFHATYVAKASGAYAKAKAA  
FEKERSNENQKRVKDYKWLKGLAEFRNOESRRARERLQDTLYPEVSEVERLER  
QTKRYVNLNLTADIEKQKHCVRQDEHYKVEVKNKAEYRENGKVLASAEVSCIQ  
RLIEDPLWESKLTIAEESVPEMKFDPAEKLGKLVSDVTNRLIIECEDAEHYRIE  
ETEMDLRWELPLFLPMKNTPEKASIOYNSCKMLAKVPEOCKESPSTYSOORLRIE  
ODLOTAYNNCOBRLQGFSDLESKVTCTRDHLREOKHFEVQGLNTINELLVNGELF  
TQARLDVATVYMEFYLOAHNIKREKVASQMAWATERREIROMAFQGVKEDLLAED  
TLKEBDDVLLDMDWLLRDERKRRORRLCNKIAA00RVKGF"  
9805. .10698  
/gene="hb4"  
9805. .10698  
/gene="hb4"  
/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB4 protein"  
/protein\_id="CAB37079.1"  
/db\_xref="GI:4455895"  
/db\_xref="SPTREMBL:Q923B1"  
/translation="MLLISGALFLTLGIPGLSAAISFGLIGLSALGVLMTSGILC  
LLVKEIPYVREEIPEGVSLAPSEEPALQAAQKTLAQLPKELDQIDIOEFACLR  
KLKDSYBSRSLINDAKKELRYEDFVVDLTSEIRPELROIYAOEGMDNLPLINGRSL  
MMTARSFLDLFHVSKRLGLPSGDVGRBGILKSKAETVARLMSIHCITHVAAVADR  
NSYAMAERAKAKALCALBESYRSITOSTRDKFLESEKAKIPWNGHTTWLNDKASGC  
AEKRLGMPRNVRNGLGKOSFG"  
10776. .12119  
/gene="hb5"  
10776. .12119  
/gene="hb5"  
/note="hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="HB5 protein"  
/protein\_id="CAB37080.1"  
/db\_xref="GI:4455896"  
/db\_xref="SPTREMBL:Q92963"  
/translation="WDEITFHELYEKTTLFLKRLHKCALANTPEKRSKKNLOAVEE  
ANARUKYRVWDYDOEFQKAGERLEKHALYPEVSISIRENKIOETRSNLEKAVEIE  
ENARCVCVEEDYMEERKEAREPENGKTLISPELSLSLEOFDGLKNFSEKLMEL  
EGHILKLOKEMATAEVENKILSDASRLIEIVEDVEKMPCKRIEIKTLRMAELPLPT  
KKAPEKASQYNSQNRHEIKQPKYKESLAVYTSKERYSIDDEDLRAVYTCQKRFQD  
DSGLSEVRAKREQDLRERIIOEETQDVLVERELCYSSRLNTECDGVSGVKEAPP  
GKTFYAOYDEIVRYRVQSRMTWTSERLREGVQCNKMLKGLSEDDVLEEYWL  
REERKNKEKRLVGTIKVATQORIOEFOPSDIVESNEKVSIMDKARFLFNEDHS"  
12262. .12987  
/gene="hb6"  
12262. .12987  
/gene="hb6"  
/note="hypothetical"

```
/codon_start=1
/transl_table=11
/product="HB6 protein"
/protein_id="CA837081.1"
/db_xref="GI:4453897"
/db_xref="SPTREMBL:092382"
/translation="MTAAPALIHVSPTPEETKFIIPKDSKSHALGITLLVILLY
CGAIVSGVIGSALIVGIGISTISLGVLFVLGILLRRELLEIAKQIAE
TFADLEKELMYIOSTREKIEKIEGSRYSNOGGLNATOKIIDLSESLISTSEFNDL
ROLPDEKIEILSGERLEFIANLRKOGHDVILNLNLADITAIYGPNNYKAVYIE
KAAVAVHEFIVLTITMARLELEFF"
/length=14585
/gene="hb7"
/organism="Homo sapiens"
/ncbi_taxonomy="9606:cellular organisms>9606:eukaryotes>9606:
mammalia>9606:primates>9606:hominidae>9606:homo"
/keywords="hypothetical"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB7 protein"
/protein_id="CA837082.1"
/db_xref="GI:4453898"
/db_xref="SPTREMBL:0966P3"
/translation="MYQENLRLLERLLYNSVOKSVADRLSEYKTKMVDPLIPWEE
DKKCAEAEKAFLEOKILIDYKSIIFMLNENDEINLNDPMSGLNVTFRKFOED
DSERNHKKVLIQKLENDYKLESESSSEANKKLISLDVLEDAKTEFLKQOE
VETRVKDLARVGGTVDPKODPEAKKKVELEAFLETFILDSIESELQOCLEDDIYKRE
OVKDLAPROLEEDIDIEAKREPAEDLSINRLKSKSTMIDRAKHLENADSTW
WTSQIDMKDKARKLIKEDITSVLEPIDIEITCLSELEPLITRELLKSLFKRI
CSFTLLKMTSVEENNIYQVEYQONLDFKILQGISORFGKQDDANLEEDVALOKK
RRELIQNEIIOGFNPMKEDFKAANDLYIRSTAEOKMFDVPCMBLFRYHEEVNKP
LLEIATNCADSYSDARKKLSLRLEDEKILQKIEKEEYQKKQKHADRSRHTYQK
LRIAELELEKKRI"
/length=14954
/gene="omp7"
/organism="Homo sapiens"
/ncbi_taxonomy="9606:cellular organisms>9606:eukaryotes>9606:
mammalia>9606:primates>9606:hominidae>9606:homo"
/keywords="putative"
/note="putative"
/codon_start=1
```

```
Query Match 100.0%; Score 3052; DB 2; Length 26920;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGATTTGCGTCGCGGATTTCTCTAGTTTCTTAAACATGCTCAGCTTC 60
DB 498 ATGGGATTTGCGTCGCGGATTTCTCTAGTTTCTTAAACATGCTCAGCTTC 557
QY 61 GACACTCTTGAAGTCTAGCATTTCTTAAACCCAGAGATGTTTCAATGAGAT 120
DB 558 GACACTCTTGAAGTCTAGCATTTCTTAAACCCAGAGATGTTTCAATGAGAT 617
QY 121 AGTCAGATSCAGAGCTTCTTAATATGTTCAAGCTGGGATGTCATAGCTTACGT 180
DB 618 AGTCAGATSCAGAGCTTCTTAATATGTTCAAGCTGGGATGTCATAGCTTACGT 677
QY 181 GATGTCATATCTAATCGATGATGATGATGATGATGATGATGATGATGATGATG 240
DB 678 GATGTCATATCTAATCGATGATGATGATGATGATGATGATGATGATGATGATG 737
QY 241 TCAGGAGTGTGAGTGTGAGGAGAAATCATGAGGTTATATTTAATATATTTCTCA 300
DB 738 TCAGGAGTGTGAGTGTGAGGAGAAATCATGAGGTTATATTTAATATATTTCTCA 797
QY 301 GGAACATCAAGAGAGGGGCTGACTTTGTTGCCAAGATCCTCAAGACAGGACGTTT 360
DB 798 GGAACATCAAGAGAGGGGCTGACTTTGTTGCCAAGATCCTCAAGACAGGACGTTT 857
QY 361 TCGGGGTTCCACGCTCTCTTTATTCAGAGCCCGGAAATATTAAGAAGAGGATGT 420
DB 858 TCGGGGTTCCACGCTCTCTTTATTCAGAGCCCGGAAATATTAAGAAGAGGATGT 917
QY 421 CTCATATCAAAAAATGCACTTATGCTTAAACATATATGATGCTTTTGAACAAAC 480
```

```
DB 918 CTCATATCAAAAAATGCACTTATGCTTAAACATATATGATGCTTTTGAACAAAC 977
QY 481 CAAAGTAGACTTAAGGCGGAGCTATATAGTGGGCGGANTGTACTATATAGTGAACACTAC 540
DB 978 CAAAGTAGACTTAAGGCGGAGCTATATAGTGGGCGGANTGTACTATATAGTGAACACTAC 1037
QY 541 GATCCGCTCTCTTCTTCAAGAAATGACAGCCATTTGGAGTGTCTATCCATCTTTCAGGT 600
DB 1038 GATCCGCTCTCTTCTTCAAGAAATGACAGCCATTTGGAGTGTCTATCCATCTTTCAGGT 1097
QY 601 CCCCTACAGATTTGAGTAAATCAGGACAGAGATAGATTTGACCAAAATCTGCCAAGAT 660
DB 1098 CCCCTACAGATTTGAGTAAATCAGGACAGAGATAGATTTGACCAAAATCTGCCAAGAT 1157
QY 661 GGTTCGAGAGGGGCTTGTACTCCGATGGGATATATATATATATATATATATATATAT 720
DB 1158 GGTTCGAGAGGGGCTTGTACTCCGATGGGATATATATATATATATATATATATATAT 1217
QY 721 CTATTTGAGAAATGAGGACATTTGACTACTGCTATAGTAAAGGAGGGGCTGTCTGTGT 780
DB 1218 CTATTTGAGAAATGAGGACATTTGACTACTGCTATAGTAAAGGAGGGGCTGTCTGTGT 1277
QY 781 CTTCACACTTCAGGAGTACTACTGCTACTGCTATATATATATATATATATATATATAT 840
DB 1278 CTTCACACTTCAGGAGTACTACTGCTACTGCTATATATATATATATATATATATATAT 1337
QY 841 TTAGCTTTGAGAAACCAATTCATATATATATATATATATATATATATATATATATAT 900
DB 1338 TTAGCTTTGAGAAACCAATTCATATATATATATATATATATATATATATATATATAT 1397
QY 901 AGCATCTCTTCAGAGGCTCTACTGCTATATATATATATATATATATATATATATATAT 960
DB 1398 AGCATCTCTTCAGAGGCTCTACTGCTATATATATATATATATATATATATATATATAT 1457
QY 961 AATTTAGTGAGCTATTTGCCATTTGATATGAGAGGAGATCACTTATATATATATATAT 1020
DB 1458 AATTTAGTGAGCTATTTGCCATTTGATATGAGAGGAGATCACTTATATATATATATAT 1517
QY 1021 GGAACATTTATCTTCAAGGAAACCGGACGATTTACCGTTTGGATGGATCCATCTT 1080
DB 1518 GGAACATTTATCTTCAAGGAAACCGGACGATTTACCGTTTGGATGGATCCATCTT 1577
QY 1081 TTTACAAATGCTAAATCTCGAAATTTACAGGCGAAGATGATGCTCTATATATATATAT 1140
DB 1578 TTTACAAATGCTAAATCTCGAAATTTACAGGCGAAGATGATGCTCTATATATATATAT 1637
QY 1141 GATCCTATTAATCTTGAAGCAGATGGGCTTACCAATTTGAATATCAACGAGATCCTTAA 1200
DB 1638 GATCCTATTAATCTTGAAGCAGATGGGCTTACCAATTTGAATATCAACGAGATCCTTAA 1697
QY 1201 AATTAAGATACAGGAGGACCATCTCTTTTCTGAGAAAGAGTCTAGCAACGATCCT 1260
DB 1698 AATTAAGATACAGGAGGACCATCTCTTTTCTGAGAAAGAGTCTAGCAACGATCCT 1757
QY 1261 AGGATTTTAAATCTTAAATCCCTCAAGCTCAACTGCTCTGAGGATCTTATATATAT 1320
DB 1758 AGGATTTTAAATCTTAAATCCCTCAAGCTCAACTGCTCTGAGGATCTTATATATAT 1817
QY 1321 AAAGAGGGGCGGAGTCAAGTTCAGGAGTTCAGGAGTTCAGGAGTTCAGGAGTTCAGG 1380
DB 1818 AAAGAGGGGCGGAGTCAAGTTCAGGAGTTCAGGAGTTCAGGAGTTCAGGAGTTCAGG 1877
QY 1381 TTAGATTTAGGAAACCAATCTGATAGCTCTTAAAGAGACATTTGCCATCAGAGGCTGCG 1440
DB 1878 TTAGATTTAGGAAACCAATCTGATAGCTCTTAAAGAGACATTTGCCATCAGAGGCTGCG 1937
QY 1441 ATAGATTAATATAGCTTAAAGCTCATCTCAAGAGGAGCTCTTAAAGAGACATCTTAA 1500
DB 1938 ATAGATTAATATAGCTTAAAGCTCATCTCAAGAGGAGCTCTTAAAGAGACATCTTAA 1997
QY 1501 AATTAACAGATATCCGAGAGGACTCTATAGAACTTATCTGAGCTTATGCAATGCTAT 1560
DB 1998 AATTAACAGATATCCGAGAGGACTCTATAGAACTTATCTGAGCTTATGCAATGCTAT 2057
```

OY	1561	GAAATCTCAGAAATGACAAATTACACAGAGTTCCTCCCTCTCTCTCTTTTAGAGCGCTGGACC	1620
Db	2058	GAAATCTCACAATAAGAATAAATTCACAGAGCTTCCTCCCTCTCTCTCTTTAGAGCGCTGGACC	2117
OY	1621	GGGGGTAGTGACTGTAACTGCTGGAGAGATTTCTCAACCGTAAGTCCCATTATGTTTT	1680
Db	2118	GGGGGTAGTGACTGTAACTGCTGGAGAGATTTCTCAACCGTAAGTCCCATTATGTTTT	2177
OY	1661	CAGAAGCAATTTGGAATTTAGCTTGGACAGCACTGGAAACAAAGTTGGAGAATTCCTTGG	1740
Db	2178	CAGAAGCAATTTGGAATTTAGCTTGGACAGCACTGGAAACAAAGTTGGAGAATTCCTTGG	2237
OY	1741	GATTAATAATTAATTTAAAGCCTAGACCGCTGAAAAAGAAATTTTGGTTCCTAAATCTTG	1800
Db	2228	GATTAATAATTAATTTAAAGCCTAGACCGCTGAAAAAGAAATTTTGGTTCCTAAATCTTG	2297
OY	1801	TGGGGGAATGCTGTAAATGTGCAGATCCTTAATGCAAGTTCAAGAGACCAGCATCAGAC	1860
Db	2298	TGGGGGAATGCTGTAAATGTGCAGATCCTTAATGCAAGTTCAAGAGACCAGCATCAGAC	2357
OY	1861	TTAACAGACAGTTCGAGGGCTGTGATCGATGGAAATTTGGSAATTTCTCCATGATATCGC	1920
Db	2358	TTAACAGACAGTTCGAGGGCTGTGATCGATGGAAATTTGGSAATTTCTCCATGATATCGC	2417
OY	1921	TCCAGACAGCATATNAAGTACCGCTCATACAGCGGTGGATGTGTTCTATCTGTAATATAT	1980
Db	2418	TCCAGACAGCATATNAAGTACCGCTCATACAGCGGTGGATGTGTTCTATCTGTAATATAT	2477
OY	1961	GAGATCACACCTTAAGCACTTACTTGATGGCATTTTCCCAAACCTCTTAAGTAGACAGAG	2040
Db	2478	GAGATCACACCTTAAGCACTTACTTGATGGCATTTTCCCAAACCTCTTAAGTAGACAGAG	2537
OY	2041	GACATGAGGGTTTCCAACAACAGATACAGATATATTAGATGCGATGTCATCAATAT	2100
Db	2538	GACATGAGGGTTTCCAACAACAGATACAGATATATTAGATGCGATGTCATCAATAT	2597
OY	2101	AACAACCTCCAGGGAATATTTTCCGTATGCTTGGCGCTAACCCCTAATGTAAACGTGGG	2160
Db	2598	AACAACCTCCAGGGAATATTTTCCGTATGCTTGGCGCTAACCCCTAATGTAAACGTGGG	2657
OY	2161	ATTCCTCAAGAAAGGTTCTTCAAAANCCCTAATGATTTTTCATTTTGTGTCCTAT	2220
Db	2658	ATTCCTCAAGAAAGGTTCTTCAAAANCCCTAATGATTTTTCATTTTGTGTCCTAT	2717
OY	2221	GGTCATGCCACCAATGATGAAAACAGCACTACGCAAAATTTCCCTATGGTGAANAACAGC	2280
Db	2718	GGTCATGCCACCAATGATGAAAACAGCACTACGCAAAATTTCCCTATGGTGAANAACAGC	2777
OY	2281	TGGAGAAACCAATTTGTGGCTATNAGATGCGGAGGAGAGCATGCCCTCTATTTGATAG	2340
Db	2778	TGGAGAAACCAATTTGTGGCTATNAGATGCGGAGGAGAGCATGCCCTCTATTTGATAG	2837
OY	2341	AACGGAAGACTTTTCCAAAGTGGCCATCCCATTTATGAACCTCAATTTAGTTATGCTAT	2400
Db	2838	AACGGAAGACTTTTCCAAAGTGGCCATCCCATTTATGAACCTCAATTTAGTTATGCTAT	2897
OY	2401	CAGGAGATTTCAAAGAGAGCACTGACAGATGGCCGTAGATTTAGTAATGGAGTTTAA	2460
Db	2898	CAGGAGATTTCAAAGAGAGCACTGACAGATGGCCGTAGATTTAGTAATGGAGTTTAA	2957
OY	2461	TTCGATTTCTGTAACCTTAGGCATAGCCTTTGGAGAACTGGCACTTTCAGAGATGTA	2520
Db	2958	TTCGATTTCTGTAACCTTAGGCATAGCCTTTGGAGAACTGGCACTTTCAGAGATGTA	3017
OY	2521	TATACACTTATGTTTCCTCTATTTCTGTAATTTTCCGTAAGATCCCTCATGTAAC	2580
Db	3018	TATACACTTATGTTTCCTCTATTTCTGTAATTTTCCGTAAGATCCCTCATGTAAC	3077
OY	2581	GCTCTGTGATTAAGCGAGCACTCTGGCTTGTTCGGCGACAGACATCAAGACATGCT	2640
Db	3078	GCTCTGTGATTAAGCGAGCACTCTGGCTTGTTCGGCGACAGACATCAAGACATGCT	3137

QY	2641	TTTGTAGGAGTGGAAACGGGTCCGATACCTTTAAACGATATACAGCTCTTATGTGCA	2700
Db	3138	TTTGTAGGAGATGGAAACGGGTCCGATACCTTTAAACGATATACAGCTCTTATGTGCA	3197
QY	2701	GGAGATATAGAAATCCGCCCCCATGTCAGCAATTAATATATAACCTGAGAGCAAAATTT	2760
Db	3198	GGAGATATAGAAATCCGCCCCCATGCTAGCAATTAATATATAACTGTGGAAGCAAAATTT	3257
QY	2751	CGTTTTAGAAAGTTTCCATTTGCCGTGTGTGTTCCGGATCTTAACATATAATCCGTAGCT	2820
Db	3258	CGTTTTAGAAAGTTTCCATTTGCCGTGTGTGTTCCGGATCTTAACATATAATCCGTAGCT	3317
QY	2821	ATGCATCATAGACATTTGGGTTTCTCCAACTGTGTGTGGAAATATACGACATTTATATGCA	2880
Db	3318	ATGCATCATAGACATTTGGGTTTCTCCAACTGTGTGTGGAAATATACGACATTTATATGCA	3377
QY	2881	TAAAGCAATACCTCATACCTACCTCAGCCCCAGAGACATCTTAAGGGCTCTTAATTTGT	2940
Db	3378	TAAAGCAATACCTCATACCTACCTCAGCCCCCTAGAGACATCTTAAGGGCTCTTAATTTGT	3437
QY	2941	CTAACTTCGTATTTTATTCGAGAAATCCTTTACGTTCTTGTGGTGTCTTCCAGAGAGT	3000
Db	3438	CTAACTTCGTATTTTATTCGAGAAATCCTTTACGTTCTTGTGGTGTCTTCCAGAGAGT	3497
QY	3001	TCTCTAACGAATCATAGGAAATCCAGGGTCTGTCTTGAATCCTTTGGCA	3052
Db	3498	TCTCTAACGAATCATAGGAAATCCAGGGTCTGTCTTGAATCCTTTGGCA	3549

RESULT#	3				
LOCUS	AE001585	10766 bp	DNA	BCF1	01-DEC-2000
DEFINITION	Chlamydia pneumoniae section 1 of 103 of the complete genome.				
ACCESSION	AE001585	AE001363			
VERSION	AE001585.1	GI:4376255			
KEYWORDS					
SOURCE					
ORGANISM	Chlamydophila pneumoniae CWL029. Chlamydophila pneumoniae CWL029 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.				
REFERENCE	1 (bases 1 to 10766) Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W., Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S. Comparative genomes of Chlamydia pneumoniae and C. trachomatis Nat. Genet. 21 (4), 385-389 (1999)				
TITLE	99206606 10192388 2 (bases 1 to 10766) Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S. Direct Submission Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA				
JOURNAL	Location/Qualifiers				
FEATURES	1..10766 <code>/organism="Chlamydophila pneumoniae CWL029"</code> <code>/strain="CWL029"</code> <code>/db_xref="taxon:115713"</code> <code>complement(1)..282)</code> <code>/gene="Cpn0001"</code> <code>complement(1)..282)</code> <code>/gene="Cpn0001"</code> <code>codon_start=1</code> <code>/transl_table=11</code> <code>/product="CT001 hypothetical protein"</code> <code>/protein_id="AAD18159.1"</code> <code>/db_xref="GI:4376256"</code> <code>/translation="MLKRIIGLSLLIYICALNWGLIGITHNNKNTITAKLCGVSTP ATQTTTIIIGAGVICTLSPFCPSKSRSHHSDSCSGCGHSHSDKN"</code> 573..878 <code>/gene="gacC"</code> <code>/note="Cpn0002"</code> 573..878 <code>/gene="gacC"</code>				
CDS					

/codon\_start=1  
/transl\_table=11  
/product="Glu tRNA Gln Amidotransferase (C subunit)"  
/protein\_id="AADI8160.1"  
/db\_xref="GI:4376257"  
/translation="MEQFLDREIILLKAKASALQISEELIOEYQTSLSAVITSMKEA  
LAIEDADSCESLTFMHVYVEDLRDVSYSDENREELKRVNPEISLGILKVPVAVIK"  
895..2373  
/gene="gata"  
/note="Cp00003"  
895..2373  
/gene="gata"  
/codon\_start=1  
/transl\_table=11  
/product="Glu tRNA Gln Amidotransferase (A subunit)"  
/protein\_id="AADI8161.1"  
/db\_xref="GI:4376258"  
/translation="MYRSALIELAKAVTLGELTATGVTOHFFHRIEAEQGVATSL  
CKRQLEQAEILDKRSRGLAGVPGIKDNIHTGKTTCASRVLENYQPPED  
AKVERIKKEDGIIIGKLNDEKAMSTLYSAFPHNPMDSRVPGSSGSAAY  
SAFCVYALGSDTGGSIROPACGVYGFPSYGAIVRGVAFASLDQICPLANTV  
EDVLMADVPSGHPDPAATREFRDSFMSKSTEVKVIQVPTFLEGLRDIKRF  
FSLAIFEBEGTHLVDELIDILSHAVSIYIIASAATNLARFDGVRGYSRQAT  
ISQYDLSRGEQGEKEMRIILGNVLSAERQNYKKATAVRAKIVKAFRAFKC  
EILAMPVCSPAFEIGELIDPVLYLQDIYVAMNLAVIPALVPSGFSKEGLPLGIQ  
IIGQGGQDQOQVGVSGFQEHQIKOLFQKRAKAVVIGGS"  
2370..3836  
/gene="gata"  
/note="Cp00004"  
2370..3836  
/gene="gata"  
/codon\_start=1  
/transl\_table=11  
/product="(pet112) Glu tRNA Gln Amidotransferase (B  
subunit)"  
/protein\_id="AADI8162.1"  
/db\_xref="GI:4376259"  
/translation="MSAVYADMEVIGLEHVHVELNASKLFSSALNRFGDEPTNIST  
VCGILGSLPYLNQSAVEKAVYLEAQTIEDGALKHGEFAGVDYNAGVPLEITVSKP  
PIIRGRIRAVQGEERFELAQTHIEDGALKHGEFAGVDYNAGVPLEITVSKP  
CMCEPDAAVAATSLVSLDYIGISDCNMEGSRVDVNVSPKSGPELRNKEIKN  
MNSFAMAOLAEKOROIDEYLNQPNKDKPLVIPATYRMDPEKKTVMIRKESAE  
DYRPEPDLPTOLTESYIERIKTLPETLYRQYIOEYLSIDIASILSDI  
APFEVACDKCKFRSLSNMYRFGRCCKTCVKPSSGIPPEGVQALVNAIDQVI  
TGTAKELADLMESPGKPEELIKKPELPSDGELEKIIIAEVLANPESTIVYK  
NGYTKALGFLVGIKRTAGKAPKRVNELLLELDKG"  
4127..6895  
/gene="pmp\_1"  
/note="Cp00005"  
4127..6895  
/gene="pmp\_1"  
/codon\_start=1  
/transl\_table=11  
/product="Polymorphic Outer Membrane Protein G Family"  
/protein\_id="AADI8163.1"  
/db\_xref="GI:4376260"  
/translation="KRFSLCGPPLVFSFLLSFTDLSATTISLPEDSFHDSQNA  
EKSYNQAGDVSLTGDVSTSNVDSNLKACQNVISGVTENGHNHGLYFNISIGT  
TREGAVLCCDDPATARFSGFSLFISQPGDKREGCLYSKALMLNNVYVRFQNT  
OSKTKGALISGANVITLVGYNDYSVFQENATREGAIIHSGPLQIVANOELIRAFQNTA  
KNSGGALISDGDIDIDONAVYLFRENALTTAIGGAVVCCPTGSSPVPVTVES  
DKNOALFEHNHISMGGAIIYARKLISISSGPTLFINNISYANSONAGGAIADTGGEI  
SLSABKGTTFPOGNRTSLPFLINGLHLQNAKFLKQARNGYSIEFPDTPTEADSGSTO  
LNIQDPKKEKTTGTLTSEGESLAPDPDFKSTIPQNVLSAGVLYIEGAVIVTSK  
FTQSPGSHLVLDLTKLILASKEDIALTTGLAIDSLSSSTAAVIAIANANKOISVTD  
SIELSPHNAVEDLNRNSQFPLSLSPGAGSVTVAGFLPVPKIRGYGGMKIL  
AMTGKNKYGEFEMDKINYPKPEKGNLYPNILKMNADVSLIMVOEHHSLSLTD  
RGIMDIGNFEHVSASEDNIRRHNSGIVLVSNNEITPKHTSAFQOLSRDQY  
AYSNNRYRYLGSYLYQYTSIGNIRYIRSRNPVNVAGILSRFLQNPIMIRFLCAY  
GHATDMKTDYANFPAVKNKSNRNKCAIECGSGMPLVLENGSLFGAIVPEKQLQVY  
ATQGFKEITADRRPNSGLTSLVPLDIRKELASDVLYIDFESFYPIDFPRDP  
SCFAALVIGSDSWLVPAHAHVSRHAFVSGTGRHFNDEYITELLRCGSIIECRPHARNYNI  
NCGSKFRF"

gene complement(7138..7293)  
/gene="Cp00006"  
CDS complement(7138..7293)  
/gene="Cp00006"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AADI8164.1"  
/db\_xref="GI:4376261"  
/translation="MOEPLRSALLERLSEMLVLLGVSPSPETRTSTPEKDAQNLPKDSR  
NRTLESL"  
7605..10499  
/gene="Cp00007"  
7605..10499  
/gene="Cp00007"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AADI8165.1"  
/db\_xref="GI:4376262"  
/translation="MKRLIAVLIALSTIILAGVYLVTAIPGLSVISSPAGNG  
ACALGCVMLAGIDVILKREVPVILASVTTPTGSPRGISISGADSTIRLPLYL  
LDGHPQSMRKLRIIAVILVFSIIILASGVILTYAIPGLSVISSPAGMGALGC  
VMALGIDVILKREVPVILASVTTPTGSPRGISISGADSTIRLPLYLDEGHP  
QSMRKLRIIAVILVFSIIILASGVILTYAIPGLSVISSPAGMGALGVYALG  
IDVILKREVPVILASVTTPTGSPRGISISGADSTIRLPLYLDEGHPQSMRKL  
VESHLNEMKSLPYDGHGLEKTKHQIRVANSISLKAQPEFLDIRIFEEFEPLSKR  
KRILDIATVTVLVERKILTEQLERNNLKRAFSYLYQDSYFKIIONFEKLAKEMILSKS  
ICRFTTIFENHEHVAKSLAKNAVILKRYLSLQSYDIDGMSKAKMLHGNPF  
SLEDNKTITKKEHAEMLESLSYRKVFLASDENVVDTPDKMIDLSIGICRDLASE  
ISRDEOMOKRAHLKHOESLYTOARDRLTDSKENVKDELEKAOEYISMERVKKEPI  
ERYOERIRAIQKLYPNILREERTGQEVTPVOCSTTASDILGIRTEVSREDN  
QNOESCVRKLSHEVMSWEVKQDYGPKKKEFPDQMSLERFTEHIEELVLYQKYS  
KHLSTYKRVNKKVQYAKFRKLVLSDEGLAQYESASLLTOELPLIATRGALAE  
KAYFKSLCOALSKAKPYFEDPREDSDPTQIRALTLRQEKAKASLEEDIKRFSME  
NDIAEERRLKESKQTERFAGLGLVLEIAVESYDLSLNTMEGTPESEKVFYSMWL  
NYNEKRRKRTKLVETQRYRDFKNALEAMQNEERLALQEEISIQAPSS"  
BASE COUNT 3165 a 2012 c 2392 g 3197 t  
ORIGIN  
Query Match 99.8%; Score 3047.2; DB 1; Length 10766;  
Best local Similarity 99.9%; Pred. No. 0;  
Matches 3049; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGGGATTTTCGCTCGCGGATTTTCCTAGTTTTCTTTAACTTGCCTCAGCTTC 60  
|||||  
Db 4127 ATGGGATTTTCGCTCGCGGATTTTCCTAGTTTTCTTTAACTTGCCTCAGCTTC 4186  
QY 61 GACACTCTTTGAGTCTACTACAGATTTCTTTAACCCAGAAAGATTTTCAAGAGAT 120  
|||||  
Db 4187 GACACTCTTTGAGTCTACTACAGATTTCTTTAACCCAGAAAGATTTTCAAGAGAT 4246  
QY 121 AGTCGAGATCGAGAACCTTCTTATATGTTCAAGCTGGGAGTGTCTATAGCCTTACTGAT 180  
|||||  
Db 4247 AGTCGAGATCGAGAACCTTCTTATATGTTCAAGCTGGGAGTGTCTATAGCCTTACTGAT 4306  
QY 181 GATGTCATATCTAATCACTGATTAAGTCTGCAATTAATAAAGCTGCTCAATGTGAC 240  
|||||  
Db 4307 GATGTCATATCTAATCACTGATTAAGTCTGCAATTAATAAAGCTGCTCAATGTGAC 4366  
QY 241 TCGGAGTGTGACGCTGCGAGAAATCATGAGGTATATATTTTAAATATTTCTGCTCA 300  
|||||  
Db 4367 TCGGAGTGTGACGCTGCGAGAAATCATGAGGTATATATTTTAAATATTTCTGCTCA 4426  
QY 301 GGAACCTCAAGAGAGGCGGTACTTGTGTGCAAGATCCTCAAGCAAGCGACGTTT 360  
|||||  
Db 4427 GGAACCTCAAGAGAGGCGGTACTTGTGTGCAAGATCCTCAAGCAAGCGACGTTT 4486  
QY 361 TCTGGTGTCTCCACGCTCTCTTTTATTCAGAGCCCGGAGATATTTAAAGAACAGGATGT 420  
|||||  
Db 4487 TCTGGTGTCTCCACGCTCTCTTTTATTCAGAGCCCGGAGATATTTAAAGAACAGGATGT 4546



QY 421 CTGATTCAGAAAAATGCACTTATGCTCTTAAACAATATGATGCGTTTGAACAAAC 480  
| | | | |  
Db 4547 CTCATTCAGAAAAATGCACTTATGCTCTTAAACAATATGATGCGTTTGAACAAAC 4606  
QY 481 CAAAGTAAAGCTAAAGCGGAGCTATTATGCGGCGAATGTTACTATATGAGCACTAC 540  
| | | | |  
Db 4607 CAAAGTAAAGCTAAAGCGGAGCTATTATGCGGCGAATGTTACTATATGAGCACTAC 4666  
QY 541 GATTCGCTCTCTTCTATCGAAGATGCAAGCACTTTGGAGGTGTATCCATTCCTCAGT 600  
| | | | |  
Db 4667 GATTCGCTCTCTTCTATCGAAGATGCAAGCACTTTGGAGGTGTATCCATTCCTCAGT 4726  
QY 601 CCCCTACAGATTGCACTAAATCAAGCGAGATAGATTAAGATTTCACAAAATACTGCCAAGAT 660  
| | | | |  
Db 4727 CCCCTACAGATTGCACTAAATCAAGCGAGATAGATTAAGATTTCACAAAATACTGCCAAGAT 4786  
QY 661 GGTTCGAGAGGCGCTTGTACTCGAGTGTGATTTATATGATCGAATGCTTATGTT 720  
| | | | |  
Db 4787 GGTTCGAGAGGCGCTTGTACTCGAGTGTGATTTATATGATCGAATGCTTATGTT 4846  
QY 721 CTATTCGAGAAAATGAGGCACTGACTGCTATAGTAAAGGAGGCGCTGTCTTGT 780  
| | | | |  
Db 4847 CTATTCGAGAAAATGAGGCACTGACTGCTATAGTAAAGGAGGCGCTGTCTTGT 4906  
QY 781 CTTCGCCACTTCAGGAAAGTACTCTCAGTTCTATTCGACTTCTCTGCAATTAACAG 840  
| | | | |  
Db 4907 CTTCGCCACTTCAGGAAAGTACTCTCAGTTCTATTCGACTTCTCTGCAATTAACAG 4966  
QY 841 TTTAGTCTTTAAAGAAACCATTCATATAGGTGGCGGAGCCATTTATGCTAGAAACTT 900  
| | | | |  
Db 4967 TTTAGTCTTTAAAGAAACCATTCATATAGGTGGCGGAGCCATTTATGCTAGAAACTT 5026  
QY 901 AGCATCTCTTCAGAGAGGTCTCTACTCTATTTATCAATATATATATATGCAAAATTCGCA 960  
| | | | |  
Db 5027 AGCATCTCTTCAGAGAGGTCTCTACTCTATTTATCAATATATATATATGCAAAATTCGCA 5086  
QY 961 AATTTAGGTGAGCTATTTGCAATGATCTGAGAGGAGATCACTTTATACAGCAGAGAAA 1020  
| | | | |  
Db 5087 AATTTAGGTGAGCTATTTGCAATGATCTGAGAGGAGATCACTTTATATCAGCAGAGAAA 5146  
QY 1021 GGAACATTTACATTCAGAGAAACCGGACGATTAACCGTTTGGATGGCATTCATCTT 1080  
| | | | |  
Db 5147 GGAACATTTACATTCAGAGAAACCGGACGATTAACCGTTTGGATGGCATTCATCTT 5206  
QY 1081 TTTCAAAAATGCTAAATCTCGAAATTTACAGCGAGAAATGATGCTATAGAAATTTAT 1140  
| | | | |  
Db 5207 TTTCAAAAATGCTAAATCTCGAAATTTACAGCGAGAAATGATGCTATAGAAATTTAT 5266  
QY 1141 GATTCCTATTACTCTGAGCAGATGGGTCTACCAATTTGAATATCAAGAGATCTTAA 1200  
| | | | |  
Db 5267 GATTCCTATTACTCTGAGCAGATGGGTCTACCAATTTGAATATCAAGAGATCTTAA 5326  
QY 1201 AATTAAGAGTACAGAGGACATACCTTTCTGAGAAAGAGTCTAGAAACGATCTCT 1260  
| | | | |  
Db 5327 AATTAAGAGTACAGAGGACATACCTTTCTGAGAAAGAGTCTAGAAACGATCTCT 5386  
QY 1261 AGGAGATTTTAAATCTCAATCTCTCAGAAAGTCAACTGTCTGAGAGTACTTATGTT 1320  
| | | | |  
Db 5387 AGGAGATTTTAAATCTCAATCTCTCAGAAAGTCAACTGTCTGAGAGTACTTATGTT 5446  
QY 1321 AAAGAGGGGGCGGAGTCAAGTTTCAAAATTCACGCACTCTCCAGAGTGCATTTAGTT 1380  
| | | | |  
Db 5447 AAAGAGGGGGCGGAGTCAAGTTTCAAAATTCACGCACTCTCCAGAGTGCATTTAGTT 5506  
QY 1381 TTTGATTTTAAAGAACCAATCGATAGCTCTTAAGAGAACATTTGCCATCAAGGCTTGGC 1440  
| | | | |  
Db 5507 TTTGATTTTAAAGAACCAATCGATAGCTCTTAAGAGAACATTTGCCATCAAGGCTTGGC 5566  
QY 1441 ATAGATATAGATGCTTAAAGCTCATCTCAACAGCAGCTCTTATTAAGCAAAACACCGCA 1500  
| | | | |  
Db 5567 ATAGATATAGATGCTTAAAGCTCATCTCAACAGCAGCTCTTATTAAGCAAAACACCGCA 5626  
QY 1501 AATTAACAGATATCGTGAGGAGCTATATAGAACTTATCTGCGCTACTGGCAATGCCAT 1560  
| | | | |

Db 5627 AATTAACAGATATCGTGAGGAGCTATATAGAACTTATCTGCGCTACTGGCAATGCCAT 5686  
| | | | |  
QY 1561 GAAGATCTCAGAAATGAGAAATTTCAACAGACGTCCCTCTCTCTTTAGAGCCCTGGAGCC 1620  
| | | | |  
Db 5687 GAAGATCTCAGAAATGAGAAATTTCAACAGACGTCCCTCTCTCTTTAGAGCCCTGGAGCC 5746  
QY 1621 GGGGAGTGTGACTGATTAAGTGTGAGATTTCCATCCGTAAGTCCCATATATGTTTT 1680  
| | | | |  
Db 5747 GGGGAGTGTGACTGATTAAGTGTGAGATTTCCATCCGTAAGTCCCATATATGTTTT 5806  
QY 1681 CAAAGCAATTTGGAATTTAGCTTGACAGGAACTGGAACAAAGTTGGAGAAATCTCTG 1740  
| | | | |  
Db 5807 CAAAGCAATTTGGAATTTAGCTTGACAGGAACTGGAACAAAGTTGGAGAAATCTCTG 5866  
QY 1741 GATTAATTAATTAATTAAGCTTACGCTGAGAAAGAAAGAAATTTAGTTCCATATCTTG 1800  
| | | | |  
Db 5867 GATTAATTAATTAATTAAGCTTACGCTGAGAAAGAAAGAAATTTAGTTCCATATCTTG 5926  
QY 1801 TGGGGAATGCTGTAATGTCAGATCCTTAATGACAGTTCAAGAGACCAATGCATCGAGC 1860  
| | | | |  
Db 5927 TGGGGAATGCTGTAATGTCAGATCCTTAATGACAGTTCAAGAGACCAATGCATCGAGC 5986  
QY 1861 TTTACAGACAGATCGAGGCGTGTGGATGATGGAATTTGGAAATTTCTTCATATCTGCC 1920  
| | | | |  
Db 5987 TTTACAGACAGATCGAGGCGTGTGGATGATGGAATTTGGAAATTTCTTCATATCTGCC 6046  
QY 1921 TTCCGAACAAATTAAGGTACCGCATACAGGCGTGGATATGTTATGCTATGTAATTAAT 1980  
| | | | |  
Db 6047 TTCCGAACAAATTAAGGTACCGCATACAGGCGTGGATATGTTATGCTATGTAATTAAT 6106  
QY 1981 GAGATCAACACTAAGCACTATCTTCTGATGAGCAATTTCCCACTCTTTAGTAGAGCAAG 2040  
| | | | |  
Db 6107 GAGATCAACACTAAGCACTATCTTCTGATGAGCAATTTCCCACTCTTTAGTAGAGCAAG 6166  
QY 2041 GACTATCGGCTTTCAACAAACGAATACAGATGATTTAGAGTGTATCTCTATCAATAT 2100  
| | | | |  
Db 6167 GACTATCGGCTTTCAACAAACGAATACAGATGATTTAGAGTGTATCTCTATCAATAT 6226  
QY 2101 ACAACCTCCCTAGGGAATTTTCCGTTATGCTTGGGTAAACCTATGTAACGTCGGG 2160  
| | | | |  
Db 6227 ACAACCTCCCTAGGGAATTTTCCGTTATGCTTGGGTAAACCTATGTAACGTCGGG 6286  
QY 2161 ATTCTCTCAAGAAAGTTCCTCAAAATCTCTTATGATTTTTCATTTTGTGTGCTTAT 2220  
| | | | |  
Db 6287 ATTCTCTCAAGAAAGTTCCTCAAAATCTCTTATGATTTTTCATTTTGTGTGCTTAT 6346  
QY 2221 GGTTCATGCCCAATGATATGAAAAACAGACTACGCAATTTCCCTATGCTGAAAAACAGC 2280  
| | | | |  
Db 6347 GGTTCATGCCCAATGATATGAAAAACAGACTACGCAATTTCCCTATGCTGAAAAACAGC 6406  
QY 2281 TGGAGAAACAATGTTGGGCTATAGAGTGGGAGGAGCAATGCTATGTTGATTTAG 2340  
| | | | |  
Db 6407 TGGAGAAACAATGTTGGGCTATAGAGTGGGAGGAGCAATGCTATGTTGATTTAG 6466  
QY 2341 AAGGAAAGCTTTTCCAAAGTGGCATCCATTTATGAACATCAATTAAGTTATGCTTAT 2400  
| | | | |  
Db 6467 AAGGAAAGCTTTTCCAAAGTGGCATCCATTTATGAACATCAATTAAGTTATGCTTAT 6526  
QY 2401 CAGGAGATTTCAAGAGACGACTGACAGATGCGCGTATGATTTAGTAATGGAGTTTAA 2460  
| | | | |  
Db 6527 CAGGAGATTTCAAGAGAGACGACTGACAGATGCGCGTATGATTTAGTAATGGAGTTTAA 6586  
QY 2461 TCGATTTCTACTCTCAGAGCATACGCTTTGAGAAGTGGCACCTTCTCAGAGTGTCTC 2520  
| | | | |  
Db 6587 TCGATTTCTACTCTCAGAGCATACGCTTTGAGAAGTGGCACCTTCTCAGAGTGTCTC 6646  
QY 2521 TATGACTTTAGTTTCTCTCTATTTCCGATATTTTCCGTAAGATCCCTATGTAAGCT 2580  
| | | | |  
Db 6647 TATGACTTTAGTTTCTCTCTATTTCCGATATTTTCCGTAAGATCCCTATGTAAGCT 6706  
QY 2581 GCTCTGCTGATTTAGGAGAGCTCCTGGCTTGTCCGAGACACGATATCAAGACATGCT 2640  
| | | | |







Dh 4907 CTTCACATTCAGAGAGTACTCTCAGTTCCTATGTCCTTCTCTGACAAATAACAG 4966  
Oy 841 TTACTCTTTGAAAGAACCATTCATATAGGTGGCGAGCCATTTATGTAAGAACTT 900  
Dh 4967 TTACTCTTTGAAAGAACCATTCATATAGGTGGCGAGCCATTTATGTAAGAACTT 5026  
Oy 901 AGCATCTCTTCAGAGAGTCTCTACTCTATTTATCAATATATATCATATGCAAAATTCGCAA 960  
Dh 5027 AGCATCTCTTCAGAGAGTCTCTACTCTATTTATCAATATATATCATATGCAAAATTCGCAA 5086  
Oy 961 AATTTAGTGAGAGCTATTTGCCATGTAATCTGAGAGGAGATCACTTTATCAGCAGAGAAA 1020  
Dh 5087 AATTTAGTGAGAGCTATTTGCCATGTAATCTGAGAGGAGATCACTTTATCAGCAGAGAAA 5146  
Oy 1021 GGAACATTTACATTCACAGAGAACCCGACAGCTTACCGTTTGTGAATGGCATCTCATCTT 1080  
Dh 5147 GGAACATTTACATTCACAGAGAACCCGACAGCTTACCGTTTGTGAATGGCATCTCATCTT 5206  
Oy 1081 TTACAAATAGCTAAATTCCTGAAATTTACAGAGCGAGAAATGAGATCTCTATAGAAATTTTAT 1140  
Dh 5207 TTACAAATAGCTAAATTCCTGAAATTTACAGAGCGAGAAATGAGATCTCTATAGAAATTTTAT 5266  
Oy 1141 GATCTATTTACTCTGTAAGAGATGGTCTACCCAAATGTAATCAACGAGATCTCTAAA 1200  
Dh 5267 GATCTATTTACTCTGTAAGAGATGGTCTACCCAAATGTAATCAACGAGATCTCTAAA 5326  
Oy 1201 AATTAAGATACAGAGGAGCATCTCTTTCTGAGAGAAAGATCTACGAAACGATCTT 1260  
Dh 5327 AATTAAGATACAGAGGAGCATCTCTTTCTGAGAGAAAGATCTACGAAACGATCTT 5386  
Oy 1261 AGGATTTTAAATCTACATTCCTCCAGAGCGTCAACCTGTCTGAGAGATCTTATGTTAT 1320  
Dh 5387 AGGATTTTAAATCTACATTCCTCCAGAGCGTCAACCTGTCTGAGAGATCTTATGTTAT 5446  
Oy 1321 AAAGAGGGGGCCGAGAGTCAAGTTTCAAAATTCACGAGCTCTCCAGAGTGCATTTAGTT 1380  
Dh 5447 AAAGAGGGGGCCGAGAGTCAAGTTTCAAAATTCACGAGCTCTCCAGAGTGCATTTAGTT 5506  
Oy 1381 TTAGATTTAGGAACCAACATGATAGCTCTTAAGAGAGACATTTGCCATCACAGGCTTGCG 1440  
Dh 5507 TTAGATTTAGGAACCAACATGATAGCTCTTAAGAGAGACATTTGCCATCACAGGCTTGCG 5566  
Oy 1441 ATGATATATAGATGCTTACGCTATCCTCAACAGCAGCTGTATTAAGCAAAACACGCA 1500  
Dh 5567 ATGATATATAGATGCTTACGCTATCCTCAACAGCAGCTGTATTAAGCAAAACACGCA 5626  
Oy 1501 AATTAACAGATATCCGTGAGAGACTCTATAGAACTTATCTCGCCTACTGSCAAATGCTAT 1560  
Dh 5627 AATTAACAGATATCCGTGAGAGACTCTATAGAACTTATCTCGCCTACTGSCAAATGCTAT 5686  
Oy 1561 GAAGATTCAGAAATGAGAAATTCACAGAGCTTCCCTCTCTCTTTAGAGCCTGGAGCC 1620  
Dh 5687 GAAGATTCAGAAATGAGAAATTCACAGAGCTTCCCTCTCTCTTTAGAGCCTGGAGCC 5746  
Oy 1621 GGGGATGATGACTGATGCTGCTGAGATTTTCTTACCGGTAAGTCCCAATTAATGTTTT 1680  
Dh 5747 GGGGATGATGACTGATGCTGCTGAGATTTTCTTACCGGTAAGTCCCAATTAATGTTTT 5806  
Oy 1681 CAAGCAATTTGAAATTTAGCTTGAGCAGAACTGGAACCAAGTTTGAGAAATTTCTTG 1740  
Dh 5807 CAAGCAATTTGAAATTTAGCTTGAGCAGAACTGGAACCAAGTTTGAGAAATTTCTTG 5866  
Oy 1741 GATTAATAATTAATTAAGCTTAGACCTGAAAGAAAGAAATTTAGTCTTAATATCTTG 1800  
Dh 5867 GATTAATAATTAATTAAGCTTAGACCTGAAAGAAAGAAATTTAGTCTTAATATCTTG 5926  
Oy 1801 TGGGGGAATCTGTAATCTGATCAGATCCCTTAATGAGATTCAGAGACCCATGATGAGC 1860  
Dh 5927 TGGGGGAATCTGTAATCTGATCAGATCCCTTAATGAGATTCAGAGACCCATGATGAGC 5986  
Oy 1861 TTACAGACAGATGAGGCTGTGTGATGATGAAATTTGGAATTTCTTCATGATCTGCC 1920  
Dh 5987 TTACAGACAGATGAGGCTGTGTGATGATGAAATTTGGAATTTCTTCATGATCTGCC 6046

Oy 1921 TCCGAGACATATATAGGTACCGCTATACAGCGGTGGAATGTTCTATCTGTAATAT 1980  
Dh 6047 TCCGAGACATATATAGGTACCGCTATACAGCGGTGGAATGTTCTATCTGTAATAT 6106  
Oy 1981 GAGATCACACCTTAAGCACTATCTGATGAGCATTTTCCCAACCTTTATGAGAGACAG 2040  
Dh 6107 GAGATCACACCTTAAGCACTATCTGATGAGCATTTTCCCAACCTTTATGAGAGACAG 6166  
Oy 2041 GACTATCGGTTTCCAAACAAAGAAATACAGAAATGATTTAGATGATCTATCAATAT 2100  
Dh 6167 GACTATCGGTTTCCAAACAAAGAAATACAGAAATGATTTAGATGATCTATCAATAT 6226  
Oy 2101 ACAACCTCCCTAGGGAATTTTCCGTTATGCTTCCGCTAACCCCTATGTAAGAGTGGG 2160  
Dh 6227 ACAACCTCCCTAGGGAATTTTCCGTTATGCTTCCGCTAACCCCTATGTAAGAGTGGG 6286  
Oy 2161 ATTCCTCAAGAAAGTTTCTTCAAAATCCTCTATGATTTTCAATTTTGTGCTTAT 2220  
Dh 6287 ATTCCTCAAGAAAGTTTCTTCAAAATCCTCTATGATTTTCAATTTTGTGCTTAT 6346  
Oy 2221 GGTATGCCCAATGATATGAAAAAGACTACGCAAAATTTCCCTATGCTGAAAAACAGC 2280  
Dh 6347 GGTATGCCCAATGATATGAAAAAGACTACGCAAAATTTCCCTATGCTGAAAAACAGC 6406  
Oy 2281 TGGAGAAACAATTTGTGGCTATGAGTGGCGAGAGGAGCATGCTCTATGATGATTTGAG 2340  
Dh 6407 TGGAGAAACAATTTGTGGCTATGAGTGGCGAGAGGAGCATGCTCTATGATGATTTGAG 6466  
Oy 2341 AACGAGACCTTTTCCAAAGGTGCCATCCATTTATGAAACACTAATTAATGATGCTTAT 2400  
Dh 6467 AACGAGACCTTTTCCAAAGGTGCCATCCATTTATGAAACACTAATTAATGATGCTTAT 6526  
Oy 2401 CAGGAGATTTTCAAAAGAGAGACTGAGATGCGCGTAGATTTAGTAATGGAGTTTAA 2460  
Dh 6527 CAGGAGATTTTCAAAAGAGAGACTGAGATGCGCGTAGATTTAGTAATGGAGTTTAA 6586  
Oy 2461 TCGATTTCTGACTCTAGAGCATAGCGTTTGAGAGAGCTGGACCTTCTCAGAGATGATCT 2520  
Dh 6587 TCGATTTCTGACTCTAGAGCATAGCGTTTGAGAGAGCTGGACCTTCTCAGAGATGATCT 6646  
Oy 2521 TATGATTTAGTTTCTCTTATATCTGATATTTTCCGTAAGAGATCCCTATGGAAGCT 2580  
Dh 6647 TATGATTTAGTTTCTCTTATATCTGATATTTTCCGTAAGAGATCCCTATGGAAGCT 6706  
Oy 2581 GCTCTGTGATTTACGAGAGACTCTGCTTGTTCGCGACACACGATACAGACATGCT 2640  
Dh 6707 GCTCTGTGATTTACGAGAGACTCTGCTTGTTCGCGACACACGATACAGACATGCT 6766  
Oy 2641 TTTGATAGGAGTGGAGAGGCTCGGTATACCTTTAAGCAGATATCTGAGCTTTATGCGA 2700  
Dh 6767 TTTGATAGGAGTGGAGAGGCTCGGTATACCTTTAAGCAGATATCTGAGCTTTATGCGA 6826  
Oy 2701 GGAAGTATTAATAATGCGCCGCCCATGCTAGGAATTAATTAATTAACCTGGAAGCAATTT 2760  
Dh 6827 GGAAGTATTAATAATGCGCCGCCCATGCTAGGAATTAATTAATTAACCTGGAAGCAATTT 6886  
Oy 2761 CGTTTTTGAAGGTTTCCATGCTGTGTGCTTCCGATCTTAACTATTAATCTGAGCT 2820  
Dh 6887 CGTTTTTGAAGGTTTCCATGCTGTGTGCTTCCGATCTTAACTATTAATCTGAGCT 6946  
Oy 2821 ATGATATATGAGATGAGGTTTCTCGAACTGTGTGAGAAATTAACGATTTATATGCA 2880  
Dh 6947 ATGATATATGAGATGAGGTTTCTCGAACTGTGTGAGAAATTAACGATTTATATGCA 7006  
Oy 2881 TAAAGGAATATCTGATACCTCAGCCCTTAGAGACATTTCTTTAGGGGTTCTTATTTGT 2940  
Dh 7007 TAAAGGAATATCTGATACCTCAGCCCTTAGAGACATTTCTTTAGGGGTTCTTATTTGT 7066  
Oy 2941 CTAAACTCTGATTTTATGAGAAATCTTTAAGCTTCTGTTGCTTGTCTCGAGAGAT 3000  
Dh 7067 CTAAACTCTGATTTTATGAGAAATCTTTAAGCTTCTGTTGCTTGTCTCGAGAGAT 7126

QY	3001	TCCTACGATCATAGGATTCACGGGTTCTGTCTTCCTTAGTCCTTTGGCA	3052
Db	7127	TCCTACGATCATAGGATTCACGGGTTCTGTCTTCCTTAGTCCTTTGGCA	7178
RESULT	5		
AE002237/c		AE002237	10044 bp DNA
LOCUS		Chlamydia pneumoniae AR39, section 65 of the complete genome.	30-MAY-2000
DEFINITION		Chlamydia pneumoniae AR39, section 65 of the complete genome.	
ACCESSION		AE002237	AE002161
VERSION		AE002237.2	GI:8163497
KEYWORDS			
SOURCE		Chlamydia pneumoniae AR39.	
ORGANISM		Chlamydia pneumoniae AR39.	
REFERENCE		Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.	
AUTHORS		1 (bases 1 to 10044)	
		Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,	
		White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Uterback, T.,	
		Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,	
		Bowman, C., Dodson, R., Gwin, M., Nelson, W., DeBoy, R., Kolonay, J.,	
		McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.	
		Genome sequences of Chlamydia trachomatis Morn and Chlamydia pneumoniae AR39	
TITLE		Nucleic Acids Res. 28 (6), 1397-1406 (2000)	
JOURNAL			
MEDLINE		2 (bases 1 to 10044)	
PUBMED		10684935	
REFERENCE			
AUTHORS		Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,	
		White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Uterback, T.,	
		Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,	
		Bowman, C., Dodson, R., Gwin, M., Nelson, W., DeBoy, R., Kolonay, J.,	
		McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.	
		Direct Submission	
TITLE		Submitted (01-MAR-2000) The Institute for Genomic Research, 9712	
JOURNAL		Medical Center Dr. Rockville, MD 20850, USA	
COMMENT		On Jun 1, 2000 this sequence version replaced gi:7189684.	
FEATURES		Location/Qualifiers	
source		1..10044	
		/organism="Chlamydia pneumoniae AR39"	
		/strain="AR39"	
		/db_xref="taxon:115711"	
		/note="synonym: Chlamydia pneumoniae AR39"	
gene		/complement(355..3123)	
		/gene="CP0770"	
CDS		/complement(355..3123)	
		/gene="CP0770"	
		/note="similar to GP:4376260; identified by sequence similarity; putative"	
		/codon_start=1	
		/transl_table=11	
		/product="polymorphic membrane protein G family"	
		/protein_id="AAF38570.1"	
		/db_xref="GI:7189685"	
		/translation="MKFSLGFPPLVFTLSVPTSLATTSITLPEDSFHGDSONA	
		ERSYVQADGVSLTGDVSIYNVNSALKACFNTSGSVTFAGNHGILYFNINISGT	
		TTEGAVLCCDQATAREGFSSTLSIOPGDIKRGCLYSKNALMLANNVREFON	
		OSKTEGGAISGANVTIGVDVSFYONATFGAIIHSSGLOIAVNOIEPFAONTA	
		KNGSGGALSDSDIDIDONAYVLPRENALTTAIGCGAVCCLPISGSPPIYTF	
		DNKQVLEFNHISIMGGALYARKLISISSGPTLTINNISYANSQNGAIAIDTGEI	
		SLSAEGKTTTFQNGTSLPEFLIHLLOAKFLKQARNGYSIEFDPLTSDSTG	
		LININDPKRKEVGTILFSGEKLANDPRDSTIPONAYLSAGYLVEIEGAEVYSK	
		FTQSPGSHVLIDLTGLKASKEDIAITGALIDISLSSSTAIVAKANANQISVTD	
		STELISPTGNAYEDLRMNSOPFLLSLERPGAGSVTVAGDELPIVSPVSGQWKL	
		AMTGTGNKGEFFMKINIKRPRKRGKLVPIILNGNAVDSLMQVETHASLITD	
		RLGLWDIGNEFHVASSEDNIRYRHNSGGYLVSNVETPKHYTSNAEQSLSRKDY	
		AVSNNEVRYLISLYOYTTSLGNIFRYASRNPNNVGLISRFLONPIMHFLCAV	
		GHATDMKTDYANFPMVKNMRRNCMAIECGSMPLVAFENGRFLQGAIPFKMLQV	
		AYOGPKETTAGDGRSNSGLTISVPLGIREKALXODVLVDFSESVIPIDFRPD	
		SCSEALVITSGDWMVPAHVSHAFVVGSTGRYHNDYTELLCRGSIIECRHARNT	
		NGSKFRP"	
		/complement(3414..4880)	
		/gene="CP0771"	
gene			
		complement(3414..4880)	
		/gene="CP0771"	
		/note="similar to PID:1653395; identified by sequence similarity; putative"	
		/codon_start=1	
		/transl_table=11	
		/product="glutamy1-tRNA(Gln) amidotransferase subunit B"	
		/protein_id="AAF38571.1"	
		/db_xref="GI:7189686"	
		/translation="MSAVYADMEVSYIGLEHVEHLNATSKLFSALNRFGDEPNNTST	
		VCTGLPGSLPVYNQSAKEVALFGCAVEGIELLSRFDKRSYFDPSPNPOTQEPH	
		PIIRGGRKATVQGEREYFELAQTHIEDDAGMLKRRGEFAGVDYRACGPLEIYSK	
		CMCPEDAVATSLVSLDYIGISDCNMEBSIRPDVNVSVPRGSPKRVKEIKN	
		MNSFAMAQALAEAKQROIDEYLNQPNKDPKLVIPATYRMDPEKKTVMRLKESAE	
		DYKYEPEPDLPTQLTESYIERIKTLPDPYDKYHYIOEYGLSIEDIASILISDNI	
		ATFEFVACDKCKNFRSLSNWVYVEFGRCCTGVKLPSGIFPEGVALQVNAIDQGI	
		TKRIKEIADIMESPGKRNPELIKPELLPMSDGELOKIIAEVYLANPESYIDYK	
		NKTRALGFLVQIMKRTGKAPRRKVNELLLELDKG"	
		/complement(4877..6355)	
		/gene="CP0772"	
		/complement(4877..6355)	
		/gene="CP0772"	
		/note="similar to GB:AL009126; identified by sequence similarity; putative"	
		/codon_start=1	
		/transl_table=11	
		/product="glutamy1-tRNA(Gln) amidotransferase subunit A"	
		/protein_id="AAF38572.1"	
		/db_xref="GI:7189687"	
		/translation="MRRISALELAARVLTGELTANGVYOHFFHRIEAGCGVATISL	
		CHEQALEQKLELDKRSRKEPGLKAGVPGIKDINHVGTLTTCASRLVNYOPPD	
		AAVEVERIKKEDDILLEKLMDEPFAMGSTLYLSAFPHTNPMPLSRVPGSGSAAV	
		SARPCVALGSDTGSIRQPAFCGVGKPSYGAVSRYGLAFASLSLOIPLANTY	
		EEVALMDVFSRDPKDATREDFRDSFMSKSTIEVPKYGIVPRTFELRDIDENE	
		FSSIAFEEGGTHNDVELDITLHVAIVYIILASAEANTLARPPGVGRYGRSPQNT	
		ISQIDLSRGEFRGKEMRRILLGNTYSAEQONYKRTATVPRAKYAPRTAREK	
		ETLAMPVCSFPAFIEGILDPVLTLYDIDITVYAMNATLPATVAVSGSKELPLGLQ	
		IIGQOGDDQVQVGSFQEHQIKQIFSKRYAKSVLGGGS"	
		/complement(6372..6677)	
		/gene="CP0773"	
		/complement(6372..6677)	
		/gene="CP0773"	
		/note="glutamy1-tRNA(Gln) amidotransferase subunit C"	
		/note="identified by match to TIGR protein family HMM TIGR00135"	
		/codon_start=1	
		/transl_table=11	
		/product="glutamy1-tRNA(Gln) amidotransferase subunit C,	
		putative"	
		/protein_id="AAF73703.1"	
		/db_xref="GI:8163498"	
		/translation="MEQFLDREELILLAKASALQSEELIOEYQTSLSAVITSMKEA	
		LAIETDDADSCESLFMHVYVNEDELDEDSTSPFNNEEPLRNPESLGGIVKPAVIK"	
		/complement(6681..6833)	
		/gene="CP0774"	
		/complement(6681..6833)	
		/gene="CP0774"	
		/note="hypothetical protein; identified by Glimmer2; putative"	
		/codon_start=1	
		/transl_table=11	
		/product="hypothetical protein"	
		/protein_id="AAF38573.1"	
		/db_xref="GI:7189688"	
		/translation="MKFKLNKQLPVPIRRTVQNKNFNLSLNDLSILARKHEIF	
		ISRLNG"	
		/complement(7249..7249)	
		/gene="CP0775"	
		/complement(7249..7249)	
		/gene="CP0775"	
		/note="conserved hypothetical protein; identified by Glimmer2; putative"	
		/codon_start=1	



Db 1923 AATTAAGATACACAGGACCACATCTCTTTTGGAGAAAAAGACTGACCAACGATCCT 1864  
 QY 1261 AGGATTTTAAATTCACATCCCTCAGAACGTCACCTGTCTGCAGATATCTTATTT 1320  
 Db 1863 AGGATTTTAAATTCACATCCCTCAGAACGTCACCTGTCTGCAGATATCTTATTT 1804  
 QY 1321 AAGAGGGGGCCGAGTCACAGTTTCAAAATTCAGCAGTCCTCCAGATGCAATTTAGTT 1380  
 Db 1803 AAGAGGGGGCCGAGTCACAGTTTCAAAATTCAGCAGTCCTCCAGATGCAATTTAGTT 1744  
 QY 1381 TTGATTTAGAACCAACATGATAGCTCCTTAAGAGAACATTTGCACACAGGCTCGGC 1440  
 Db 1743 TTGATTTAGAACCAACATGATAGCTCCTTAAGAGAACATTTGCACACAGGCTCGGC 1684  
 QY 1441 ATGATATATAGATAGCTTACCTCATCTCAACAGCAGCTGTTATTAAGCAAAACCCGA 1500  
 Db 1683 ATGATATATAGATAGCTTACCTCATCTCAACAGCAGCTGTTATTAAGCAAAACCCGA 1624  
 QY 1501 AATTAACAGATATCCGTGACGAGCTATATAGAACTTATCTGCGCTACTGGCAATGCTAT 1560  
 Db 1623 AATTAACAGATATCCGTGACGAGCTATATAGAACTTATCTGCGCTACTGGCAATGCTAT 1564  
 QY 1561 GAAGATCTCAAGATGAGAAATTCACAGAGCTTCCCTGCTCTTAAAGAGCTCGAGCC 1620  
 Db 1563 GAAGATCTCAAGATGAGAAATTCACAGAGCTTCCCTGCTCTTAAAGAGCTCGAGCC 1504  
 QY 1621 GGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
 Db 1503 GGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1444  
 QY 1681 CAAGGCAATGGAATATGCTTGACAGAGAACTTGGAACAAAGTTGAGAAATTTCTTGG 1740  
 Db 1443 CAAGGCAATGGAATATGCTTGACAGAGAACTTGGAACAAAGTTGAGAAATTTCTTGG 1384  
 QY 1741 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800  
 Db 1383 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1324  
 QY 1801 TGGGGGAAATCTGTAAATGTCAGATCTTAAATGACAGGTTCAAGAGACCAATGATGAGC 1860  
 Db 1323 TGGGGGAAATCTGTGTAGATGTCAGATCTTAAATGACAGGTTCAAGAGACCAATGATGAGC 1264  
 QY 1861 TTACAGACAAATGAGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920  
 Db 1263 TTACAGACAAATGAGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1204  
 QY 1921 TCCGAGACAAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
 Db 1203 TCCGAGACAAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1144  
 QY 1981 GAGATACACCTTAAGACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040  
 Db 1143 GAGATACACCTTAAGACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1084  
 QY 2041 GACTATGCGGTTCCCAACAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2100  
 Db 1083 GACTATGCGGTTCCCAACAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1024  
 QY 2101 ACAACCTCCCTAGGAAATATTTCCCTTATGCTTCCGTAACCTTAATTAATTAATTAATTAAT 2160  
 Db 1023 ACAACCTCCCTAGGAAATATTTCCCTTATGCTTCCGTAACCTTAATTAATTAATTAATTAAT 964  
 QY 2161 ATTCTCTCAAGAAAGTTTCTTCAAAATTCCTTATGATATTTTCAATTTTGTGTGCTTAT 2220  
 Db 963 ATTCTCTCAAGAAAGTTTCTTCAAAATTCCTTATGATATTTTCAATTTTGTGTGCTTAT 904  
 QY 2221 GGTCAATGCCACCAATATATTAAGAAACAGACTACGCAATTTTCCCTATGGTGAACACAGC 2280  
 Db 903 GGTCAATGCCACCAATATATTAAGAAACAGACTACGCAATTTTCCCTATGGTGAACACAGC 844  
 QY 2281 TGGAGAAACAAATTTGGCTATAGAGTGGAGGAGAGCAATGCTTATTTGATTTGAG 2340  
 Db 843 TGGAGAAACAAATTTGGCTATAGAGTGGAGGAGAGCAATGCTTATTTGATTTGAG 784

QY 2341 AACGAGACCTTTTCCAAGTGCCATCCCATTTATGAACATTAATTAATTAATTAATTAATTAAT 2400  
 Db 783 AACGAGACCTTTTCCAAGTGCCATCCCATTTATGAACATTAATTAATTAATTAATTAATTAAT 724  
 QY 2401 CAGGAGATTTTCAAGAGACAGCTGCAATGCGCGTATTAATTAATTAATTAATTAATTAATTAAT 2460  
 Db 723 CAGGAGATTTTCAAG 664  
 QY 2461 TCGATTTCTGATCTTATAGGATATGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520  
 Db 663 TCGATTTCTGATCTTATAGGATATGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604  
 QY 2521 TATGATTTTATGTTTCTTATATCTTATATCTTATATCTTATATCTTATATCTTATATCTTATAT 2580  
 Db 603 TATGATTTTATGTTTCTTATATCTTATATCTTATATCTTATATCTTATATCTTATATCTTATAT 544  
 QY 2581 GCTCTGATGATTAAG 2640  
 Db 543 GCTCTGATGATTAAG 484  
 QY 2641 TTGATTAAG 2700  
 Db 483 TTGATTAAG 424  
 QY 2701 GGAAGTATAGAAATGCCGCCGCCCATGCTAGAGAAATTAATTAATTAATTAATTAATTAATTAAT 2760  
 Db 423 GGAAGTATAGAAATGCCGCCGCCCATGCTAGAGAAATTAATTAATTAATTAATTAATTAATTAAT 364  
 QY 2761 CGTTTATGAAGTTTCCATATGCTGT 2820  
 Db 363 CGTTTATGAAGTTTCCATATGCTGT 304  
 QY 2821 ATGATCATATGAGATGAGGTTTCTGCAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2880  
 Db 303 ATGATCATATGAGATGAGGTTTCTGCAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 284  
 QY 2881 TAAAGGAATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2940  
 Db 243 TAAAGGAATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 184  
 QY 2941 CTAACTCTTATTTATTCAGAAATCTTACGTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3000  
 Db 183 CTAACTCTTATTTATTCAGAAATCTTACGTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 124  
 QY 3001 TCTCTAAGCAATATGAGAGATTCAGAGGTTCTGTCTTCTGAGTCTTGTGCA 3052  
 Db 123 TCTCTAAGCAATATGAGAGATTCAGAGGTTCTGTCTTCTGAGTCTTGTGCA 72

RESULT 6  
 LOCUS A81853 1200 bp DNA PAT 21-JAN-2000  
 DEFINITION Sequence 27 from Patent WO9858953.  
 ACCESSION A81853  
 VERSION A81853.1 GI:6731878  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1200)  
 AUTHORS Madsen, A. and Birkeland, S.  
 TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE  
 JOURNAL Patent: WO 9858953-A 27 30-DEC-1998;  
 MADSEN ANNA SOFIE (DK); BIRKELUND SVEND (DK)  
 FEATURES  
 source Location/Qualifiers  
 1..1200  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 <1..>1200  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAB69201.1"

CDS







```

/gene="CP0758"
/Note="hypothetical protein; identified by glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38560.1"
/db_xref="GI:7189673"
/translacion="MRLLSCPMLEFIACASPFGEQEMOCHNIQSLDANASSLGELE
SISRGVSCLEHREIAR"
complement(3336..6086)
/gene="CP0759"
/Note="This region contains a gene with one or more
premature stops or frameshifts, and is not the result of a
sequencing artifact; similar to GP:4455886; identified by
sequence similarity; putative; polymorphic membrane protein
G family, degenerate"
complement(6325..9110)
/gene="CP0760"
/Note="This region contains an authentic frame shift and
is not the result of a sequencing artifact; similar to
GP:4455886; identified by sequence similarity;
putative; polymorphic membrane protein G family, authentic
frameshift"
complement(9393..11918)
/gene="CP0761"
complement(9393..11918)
/gene="CP0761"
/Note="similar to GP:4376270; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"
/protein_id="AAF38561.1"
/db_xref="GI:7189674"
/translacion="MKIRIPRLISLYPTLSNLSGAATTELSASNPDTSTTS
ESSKTSATDSTNYFKDSYVIENTPKTGETOSTSCFKNDAAGLNTLGGFSTFS
NIDATTAAGAAIGSEPAANKTVLSEFSAISLSTKSTASTVAGLAINFKNSLNDND
KVLIDNFTSGDGAINCAGSLKIANNSLSFISGSSSTRGALHTKMLTSSGGELT
FQALRAAGHTIYFDPITVTGTSVADALNINSIDPNKNEYTITVESGKLEAE
AKDENRTSKLONAFKNGTVLKGDVLSANGSODANSKLIMDTLSIYANEST
ELTLEINIDSLRNKRIKLSAATKRIIRIDRPVYLAISDESPIONKFLNDSHYDG
ILELDAGKDIYISADRSIDAVQSPYGTQGRMTLMSTDKKATVSMARKOSNPFAEO
EAPLVNPLMGSEFIDVRSFQNFIEGTGAEYEKFTVAGISNVLRSGRENORFRI
VSGAVVASTRMPGGDTLSLFAQLFARDYFENITFAATVAGSLDLOHASLYSV
VSIILGEGLEIREILPYVTKLPCSEFQSLYGHDRMKTESLPEPPTSLDSTSW
GGYVWAGELGRVAVENTSGRPFQETVPEFKVAVYARODSPVELGASISDFSDSHL
YNLAIPGKLEKRPABQYHYVYVAMYSVDYRSPNKCTTTLSNDSGSKRTGSLNARO
AGIYVQSGFRLGAALFENGFEMWSSRSYNDASKIF"
BASE COUNT 3532 a 2722 c 2431 g 3441 t
ORIGIN
Query Match 8.9% Score 272.6; DB 1; Length 12127;
Best Local Similarity 48.2%; Pred. No. 2.8e-55;
Matches 1266; Conservative 0; Mismatches 1274; Indels 89; Gaps 14;
QY 207 CTTTCATTAAATAAGCGTCAATGTGACTCAGCAAGTGTGAGCTGCAGGANA 266
Db 8904 CACAGCATCAAAAAGCTGTTTAAACACACTAAGGCGATTGACTTTACAGGTA 8845
QY 267 TCATCATGGTTATATTTTAAATATATTTCTCAGCAACTACAAAGAGGGCTGTACT 326
Db 8844 CGGCAACTCTCATTTGCCAAACGGTGATGACGAGACTGTA---GCAAGGGCTGCT 8788
QY 327 TTGTTGCCAAGATCTCTCAAGCAAGCGCTTTTCTGGGTTCTCAGCGCTCTTTTAT 386
Db 8787 TAACAGCAGCGGTGATTAATCTACACAGTTTATAGGTTTCTTCGCTATCTTTAT 8728
QY 387 TCAGAGCCCCGAGATATTTAA-----GAACAGGATGCTCTATTCAAAAATGCA 438
Db 8727 TGGCTCTCTGGAAGTTGATTAACACGCGCAAGAGACCGCTTACTGCTCTACGGTAG 8686

```

```

QY 439 CTTATGCTCTTAAACATTATATGATGCTTTTGAACAAAACCAAGTAAAGGC 498
Db 8667 CTGATCTTTGACAAAATATGCACTTTGCTCTCAGCAAAAACCTTTCAACGATTAATGCG 8608
QY 499 GAGCATTAATAGTGGCGAATGTTACTATATAGTACCACTAGATTCGCTCTTCTAT 558
Db 8607 GGTGCTATCACCAGAAAACCTCTTCAATTAACGGGACATACATATGCTGCTGTTTCT 8548
QY 559 CATAATGACGCCACTTT---TGAGGGGTATCCATCTTACAGTCCGCTACAGATTGCA 615
Db 8547 GAAATATCTCTCTCAAGAAAGGCGAGCCATTACACTTCGATCCCTTACCATTTACT 8488
QY 616 GTAATACAGCAGATATGATTTGACAAAATCTCCCAAGATGTTCTGGAGGGCT 675
Db 8487 GGAACCAAGGGGAAGTCTTTTCTGACAAATCTTCTG---GATCTGAGCTGCA 8431
QY 676 TTGTACTCCGATGATATGATATGATATGATCAGATGCTTANGTCTTATTTGAGAAAT 735
Db 8430 ATTTTACAGAAAGCCCTCGGTGACTATTTCTAATTAAGCTAAAGTTCTTATTTGACAAAT 8371
QY 736 GAGG---CATTGACTCTGCTATAGTAAAGGGGGGCTGCTGTTCTCTCCACTTCA 792
Db 8370 AAGGTACAGAGGAGCGAGCTCTTACACAGGGGATATGTCAGAGAGTCTATCTGTGCT 8311
QY 793 GGAAGTACTACTCCAGTCTCTATTTGACTTCTCTGACAAATTAACAGTTAGCTTTGAA 852
Db 8310 TATTAAGCTAGTACATATCTAAGTACACCTCAGTGAATTCAGATGATCTTCTCAGC 8251
QY 853 AGAACCATTCCATTAATGGGTGGGAGCCATTATGTAAGAACTTATGATCTCTTCA 912
Db 8250 AACCAATACATCGAACACACGGGAGGATATCTAATGTAAGAAAGCTCAACATGCGCTTC 8191
QY 913 GGAGGTCTACTCTATTTTCAATATATATATATATATGCAAAATTCGCAAAATTTAGGTGA 972
Db 8190 GAGAGCTTACCTTATCTAGTGAATATGTCATATGAGGTRACAGCTCTTAAGGTGGA 8131
QY 973 GCTATTTGCCATTGATACAGGAGGAGATCAGTTATATACAGAGAGAAAGCAATTTACA 1032
Db 8130 GCAATGCTATGGAAGATAGTGGGGAATGATGATTTATCCGCGATAGTGTGATCATTTGC 8071
QY 1033 TTCCAAAGAAACCGGACGAGCTTAC-----CGTTTGAATGGCATCTTCT 1080
Db 8070 TTTTAAAGAAATACAGTCACTTCTTACTACTCTGAGCAAAATGAAGTGTATGCACTTA 8011
QY 1081 TTAACAAAAGCTAAATTTCTGAAATTTACAGGAGAAATGAGATGCTCTTATCAATTTAT 1140
Db 8010 GGAAGAGTGAAGATATGACACTTTGCTTCTGCTGCTGTAAGACCATCTTACTTCTAT 7951
QY 1141 GATCTTATTTACTTCTGAAAGCAGATGGGTCTACCAATTTGAATATACAGGAGATCTTAA 1200
Db 7950 GATCCCATTAATACAGATATCTCAACACAGTTACAGATGTCTTAAAGTAAATGAGACT 7891
QY 1201 AATTAAG-----AGTACACAGGAGCATCTTCTTTTGGAGAA----- 1240
Db 7890 CGGAGAGATTTGCACTACAAATATACAGGAAATCATCTTCAAGGAAAGATTAATCA 7831
QY 1241 -GAGATCTGCAAGAGATCTAGGGATTTTAAATCTACAAATCCGAGAGAGCTCAACCTG 1299
Db 7830 GAGACAGAGGCGGAGATTTCTTAAATCTTACTTGAAGCTCTACAGCTCAAGCACTTAACTCT 7771
QY 1300 TCTGAGGATTTACTGTTATTTAAAGAGGGGCGGAAGTCAAGTTTCAAAATTCAGCGCA 1359
Db 7770 TCAGAGGATCTCTATCTTAAACATGAGAGTCTCTCAGACATCAGCACTTCACTCA 7711
QY 1360 TCTCCAGATGCTCATTTAGTTTATGATTTAGAACCAACTATAGCTCTTAAGAGAC 1419
Db 7710 CAGGAGATTTCTGCTGGAATGAGACGTAGGAACTACTCTAAGAACCTGCT-----GAT 7657
QY 1420 ATTGCATACAGGCGCTGCGATATGATATGATATGATTAAGCTCATCTCAACAGCAGCT 1479
Db 7656 ACTACACCATTAACAAATTTGCTCATTTAATCATGATTTCTATAGACGGTGCAGAAAGAGCA 7597

```

QY	1480	GTATTATTAACGAACAACCCGCAATAATACAGATATCCGACGACGACTCATATGAACTATC	1539
Db	7596	AAATATGAAACCAACACTACGTCACAAAAATCTGACTTTATATCGAAGCATCATTTATTG	7537
QY	1540	TGCGCTCTGGCAATGCCATATGAGATCTCAGATAGAGAAATTCACAGACTTCCCTCG	1599
Db	7536	GACCCGACGGGACGTTTATATGAAATCATATTTAAGAAATCCTCACTCCTACGACATC	7477
QY	1600	CTCTCTTT---AGAGCCTGGAGCGGGGGTACTGTGACTGTAACTGCTGGAGATTTCCYA	1656
Db	7476	TTAGAGCTCAAAAGCTCTGGAAGCTGTAAACAGACACCGAGTGACTCCAGATCTATATG	7417
QY	1657	CCGGTAAGTCCCATATATGTTTTCAGGCAATTGG---AAATTAGCTGGACAGAACT	1713
Db	7416	GGTGAGAAATTCATTACGGCTATACGGGAACCTTTGGGGCCCAATTTGTTGGGGACAGGG	7357
QY	1714	GGAACCAAGTTGGAAATTTCTCTGGGTATTAATAATTAATTAATTAACCTTAGACCTGAAAA	1773
Db	7356	GCCTTTCACGACTCCAACTTTCACATGAGACTTAAACCTGGCTATATTCCTAATCCGAGCGT	7287
Db	7296	ATCGGCTCTTATGCCCTAATAGCTTATGGAATGGAATTTATATGATATATAGCTCTCTCCAT	7237
QY	1834	CAGGTTCAAGAGACCATGCACTACGAGCTTACGACAGATCGAGGCTGTGATGATGGA	1893
Db	7236	TATCTTATGAGAGCTCCAAACGAGGGTTCGAGGAGACCGGCTTTTGGTGTGCTGGA	7177
QY	1894	ATTGGGAATTTCTTCATGATATCTGCTCCGAGAACAAATATAGGTACCGTCATACAGC	1953
Db	7176	TTATCTACCTCTTCCATTAAGGATAGTACAAAACACGACGGGGTTTGGCATTTGAGT	7117
QY	1954	GGTGGATATGTTCTATCTGTAAATAATGAGATCACACGTAAAGCACTATATCTCATGGA	2013
Db	7116	GCGGTTATGTATAGAGAGAAACCTACATACTGTTGCAAGTAAATCTTATAGTGCGCA	7057
QY	2014	TTTTCCCACTCTTTTATGATAGACAGAACTATGGCGTTTCCACACGAATTAACGAATG	2073
Db	7056	TTTTTGTACGCTTTTGGAAAGATATAGACTCTTTGTAGCTTAACAAATCAAGGTACAGTC	6997
QY	2074	TATTTAGATCGATCTCATCAATTAATTAACACCTCCGTAAGGAAATTTTCCGTTATGCT	2133
Db	6996	TACGGAGAACTCTTATTTACGACACAGCAAAAC-----TATA	6957
QY	2134	TCGGTAACCTTAATGTAAACGTCGGGATTTCTCAGAAAGGTTTCTTCAAAATCTCTTT	2193
Db	6956	TCCTCTCTCTCCCAACATACGGCCCTGTGCTGTCTTATAGTTCTTACAGATATTCCTG	6897
QY	2194	ATGATTTTTCATTTTGTGTGCTTATGCTCATGCCACCAATGATATGAAACAGACTAC	2253
Db	6896	TTCTCTTTTCAGAAACCTWAGC-TACACCCCTACGATTAACGATCTGAAACCAAGATAT	6838
QY	2254	GCAAAATTTCCCTATGSGTGAAGAAAAACGCTGGACAAACATTTGTGGGCTATATGATGGGGA	2313
Db	6837	ACAACATATCTTACTGTAAAGAACCTGGGAAATGATAGTTTCGTTTGAATTTGGGT	6778
QY	2314	GGGAGCATGCTCTATGTGATTTTGGGAACGAGAAACCTTTTCAAGGTGACCTCCCATTT	2373
Db	6777	GGAAGAGCTCCGATTTGCTTAGTAAAGTGTCTATTTTGAAGCACTA---CATGCCCTTC	6721
QY	2374	ATGAAATACAAATTAGTTTATGCTTATCAGGAGATTTCAAGAGACGACTGCAGATGGC	2433
Db	6720	ATGAAATATGAGTTTGTATGACATCCAGGAAGGTTTTTAAAGAACAGGAAGAAAGCT	6661
QY	2434	CGTAGATTTTGTATTAAGGAGTTTAACATCGATTTCTGTACCTCTAGAGCATACGCTTTGAG	2493
Db	6660	CGTAGAATTTTGAAGTACCGCTTGTGTGAATCTTGGCTTACCTATTCGGGATCCGATTTGAT	6601
QY	2494	AAGCTGCACTTTCTAGAGATGACTCTATGACTTTAGTTTCTCTTATATCCGATATTT	2553
Db	6600	AAGGAATACAGACTCCAAAGATGCAAGTCAATCTTACTCTTTGGTTATATCTGTGGATCTTT	6541
QY	2554	TTTCGTAGAGATCCCTATGTGAAGCTGCTGCTGATTTAGGAGAGACTCCTGGCTGTT	2613

Db	6540	GTTCGTAGTACACCCCGCAGCTGTACGACAAACACTGCGAATTTAGCGGTATTTTGGAAACTT	6481
OY	2614	CCGGAGCAGCAGCAGTATCAAGACATGCTTTTGTAGGGAGTGGAAAGGGTCCGTATCACTT	2673
Db	6480	TTCCGTACCAATTTTGGCAGACAGACAGCTTTAGTCCCTCGTGCAGAGAAACATTTTTCGCTT	6421
OY	2674	AACGACTATAGTACGAGCTTTATGTGAGGAAGATAGATGAATCCGCCCATGCTAGCAAT	2733
Db	6420	AACCTAAATTTTGAAGCCCTTTAGCCAAATTTTCTTTGAATTTGGTGGTCAATCCGAAT	6361
OY	2734	TATATATATTAACGTGAGCAAAATTTGCTTTTACAGGTTCCATGTG	2782
Db	6360	TACAAATGTAGCTTAGGAGCAAAATACCAATTTCTAATGCTTAGCTTGG	6312
RESULT	8		
LOCUS	AE001587	16448 bp	DNA
DEFINITION	Chlamydia pneumoniae section 3 of 103 of the complete genome.		
ACCESSION	AE001587	AE001363	
VERSION	AE001587.1	GI:4376271	
KEYWORDS			
SOURCE			
ORGANISM	Chlamydia pneumoniae CML029.		
REFERENCE	Chlamydia pneumoniae CML029.		
AUTHORS	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
	1 (bases 1 to 16448)		
	Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,		
	Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.		
	Comparative genomes of Chlamydia pneumoniae and C. trachomatis		
	Nat. Genet. 21 (4), 385-389 (1999)		
TITLE	2 (bases 1 to 16448)		
JOURNAL	Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,		
MEDLINE	10192388		
PUBMED	99206606		
REFERENCE			
AUTHORS	2 (bases 1 to 16448)		
	Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,		
	Grimwood, J., Davis, R.W. and Stephens, R.S.		
	Direct Submission		
	Submitted (01-DEC-1998) Program in Infectious Diseases, University		
	of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA		
TITLE	Location/Qualifiers		
FEATURES	1..16448		
source	/organism="Chlamydia pneumoniae CML029"		
	/strain="CML029"		
	/db_xref="taxon:115713"		
misc-feature	172..2957		
	/note="Polymorphic Outer Membrane Protein G Family;		
	contains frame shift"		
gene	172..705		
	/gene="pmp_3.1"		
	/note="CpN0014"		
gene	615..2957		
	/gene="pmp_3.2"		
	/note="CpN0015"		
gene	3196..4971		
	/gene="pmp_4.1"		
	/note="CpN0016"		
misc-feature	3196..5953		
	/note="Polymorphic Outer Membrane Protein G Family;		
	contains frame shift"		
gene	4874..5953		
	/gene="pmp_4.2"		
	/note="CpN0017"		
gene	6302..7786		
	/gene="pmp_5.1"		
	/note="CpN0018"		
misc-feature	6302..9139		
	/note="Polymorphic Outer Membrane Protein G Family;		
	contains frame shift"		
gene	7787..9139		
	/gene="pmp_5.2"		
	/note="CpN0019"		
gene	complement(9380..11467)		
	/gene="CpN0020"		

```
CDS      complement(9380..11467)
        /gene="Cp00020"
        /codon_start=-1
        /transl_table=11
        /product="Predicted OMP [leader peptide: outer membrane]"
        /protein_id="AAD18173.1"
        /db_xref="GI:4376272"
        /translation="MKRCFLFLASFLVLMGSSADALTHQAVKKKNSYLSHFESGIV
        TLEDGLVNHNLRIQANVYVENVYGGSLKVAHGNNVWVRAKTLVDYLYEYEDT
        DCLTNGRPAWPFGLSGSMITLPTPTVIRKGYSTSESGPKKLCISGDIYVSD
        SLISGKTLRYCRIPILPLPPSLIMPEIKPPIINFRGGGFLGSLYGMISPLSR
        KHSTFELDFSEFKGVGMFNHCSOKOVPEVEMKSYVAHRLAIDMAEADNRRL
        HDPCFTKHNHVSSEYHLSDMEVADIEPNNFMKLKNGPTVRCCTMDYNEGYLT
        SSYKVSFQNAQELPYTLRQYPISTVNTGYLENIYEGVLFNFAFSOHYGENSS
        LRLAARPKHTVPLPIGLTSLSSSLIYGDVPEISSHQSLSAKIQLDLPFLHK
        SYTORHTEPEVPTETTERPRPLAKNDHYFESIDQAFHSLINDKAGIDPSYKTPR
        FRIHAKLTHTIILSTESKPTPPKACLSLPGKKNVSLDAEITMKKCHDHANI
        RNEWIGNDVAMTLESLSRSKSLIKCDRENFLLDVSRIDQLDLSPLSDHRLILGK
        LEVRBPCWNYRLSLYGMHRODTPNYLEYOMILGTLKEPHQLYGVYRRERDSRF
        FLKLDKPKKPF"
gene      complement(11484..13190)
        /gene="Cp00021"
        /complement(11484..13190)
        /gene="Cp00021"
        /codon_start=-1
        /transl_table=11
        /product="Predicted OMP [leader peptide]"
        /protein_id="AAD18174.1"
        /db_xref="GI:4376273"
        /translation="MGLFHLTLFGLLCSPLISLVAKPEPESVGHKILYSTOQOAL
        AYLLDALVGDHDFEFLKIGEDYLKSHSSDPQTRSTIIGAGLAAKSLADVL
        QAMETADPLOQLLVSAGHIGKTSDDLFLALASPVIRLEAAYRLANKNPTVI
        DHLSHPIRLPEIQCISAFILFLEETESDAYIRDLAKKASATPAALDIEGQO
        KEFLKRLTLTASPOOEALILYAGKIKDQOSYVNIKKOKPVDVATLAAQALI
        ALGKEBDALPYTKQALEPRPALALRLHPSIGIPLPIFLTKNSBAKALVALA
        LLELGDPTKLELYITERLVOPHYNETLASLSESGTTLQNMKRVNIYVQDQERLL
        LSTRLQEOILFLFRLPEKAYVPCITVYLLASOKYQATTAISFSLSHSEADLIL
        FOAKLPGPIIRAYADLAIVMLTKDPEKRSIHDYAKLLOETLFLVDTEOMRPS
        MPLYRQVPESTRKMLDILETLATSKSESDIRLLIQMTSGDAKNFVYLAGLLIKI
        VE"
gene      complement(13172..13762)
        /gene="maf"
        /note="Cp00022"
        /complement(13172..13762)
        /gene="maf"
        /codon_start=-1
        /transl_table=11
        /product="Maf-type protein"
        /protein_id="AAD18175.1"
        /db_xref="GI:4376274"
        /translation="MSLPVLGSSSPRRKFLIEKFPVPTVIPSNFDESKVSGDPI
        ATYOLAOKAVASELHSPDCIITIGPTIVSYGRIETKQDKADATOMKTLRNO
        THDVYTSIAVLHKGKLLGSSERSOISTIMIPDHRESYIDYGTINNGCAVYVCHGL
        ILKRYHGCYVNYQGLPIQTLKYLEELNIDLDYSL"
gene      complement(13791..15383)
        /gene="yijk"
        /note="Cp00023"
        /complement(13791..15383)
        /gene="yijk"
        /codon_start=-1
        /transl_table=11
        /product="ABC Transporter Protein ATPase"
        /protein_id="AAD18176.1"
        /db_xref="GI:4376275"
        /translation="MSIVLDKIGKSLGRILFDVDSVYVNPNCYGLGPNAGKSTL
        LKINGMTEPRGSTSLPKVGLIKRONIDSEHDTTVLDGVINGNRLMALQRNOLY
        LOEFTDAIGMELGEIEIIGEENGRADEABELLIGIIPMEMDKKAMAPIDLOF
        RYLCOALFGHEALDELPEYTHLDLYSINMGNFLKDEGVIVVSHRHLANTIT
        HIAIDVDTIITIPGVNDVMVEMTKASROEKADIKSKKKSOLKEFAFKAGSRA
        SOVOSRIRKIKLOPBLKKSNIORPYIREFSLDKSSGVVLSLEATIKQDDHYIH
        PPSLEIYGDKIGITGNGSLKTTMLKLALAGVASSGSIKLGHAICSYFQNHSDV
        LADCGEFLFEWLRNRKTGINDEIRSYGKRLFGGDAFKQIOLASGETARLLMAG
        MLEHNHNVILDEANNHLDLESVSALSWAINDYKGTAFVSHDRSLIDCCATKLLIFD
```

```
gene      KDKITFEDGTWVDTAGHKQL"
        complement(15438..16376)
        /gene="xerc"
        /note="Cp00024"
        /complement(15438..16376)
        /gene="xerc"
        /codon_start=-1
        /transl_table=11
        /product="Integrase/recombinase"
        /protein_id="AAD18177.1"
        /db_xref="GI:4376276"
        /translation="MTASISFLDYLIKWKVKSAPHTLRNYCLDNLKLTLEERNLA
        PSSPOLATEKRVSELPSELTKEHVRNYIKLLENGAKRTIKRCLSSISFPAHC
        VIOKILLENPAETIYELPRKELPSMTAQVQEVLMATDISKYLDRCLMELEYS
        SGIRISEIYAVKODELSTHLIRIGKKEKRRIIPVSNALOMIOIYLNHDKRLE
        KDOIATLNRFPGRRISTRSIDRSFOEYLRSLRSGLPHITRIHIAIHWLESMDLK
        TIOALLGHSLETTIYTVYSVKLKQIHOEHPHA"
BASE COUNT 4710 a 3330 c 3665 g 4743 t
ORIGIN
```

Query Match 8 9%; Score 271; DB 1; Length 16448;  
Best Local Similarity 48.1%; Pred. No. 6.9e-55;  
Matches 1265; Conservative 0; Mismatches 1275; Indels 89; Gaps 14;

```
QY 207 CTCGCAATTAATAAGCCGCTTCATGTGACCTGAGAAAGTGACGTTCCGACAGAA 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 CACAGCAATACAAAAGCTGTTTAACACACTAAGGCGGATTTGACTTTCACAGSTA 437

QY 267 TCATCATGGGTATATTTATATATATTTCTCAGCACTACAAAGAGGCGCTGACT 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 CGGAAGCTCTCTATTTGTCMAAGCGGTAGTACGAGGACTGTA---GCAGGGCGCTGCT 494

QY 327 TTGTTGCCAAGATCCCAACAGAGGACGCTTTTTCGGTCTCCACGCTCTCTTTAT 386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 495 TAAACAGCGGTGTAGATTAATCTACACGTTTATAGGTTTCTGCTATCTTTTAT 554

QY 387 TCAGAGCCCGGAGATATTTAA-----GAACAGGAGATCTCTATTCAAAAATGCA 438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 555 TCGCTCTCTCGGAAGTTCCATACCTACCGGCAAGAGACCGGTAGCTGTACGGGTAG 614

QY 439 CTTATGCTCTTAAACATTAATATATATATGCTTTGAAACAAAACCAAGTAAAGCTAAAGC 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 615 CTTGAGTTTACAAAATGTCAGTTTCTCTTCAGCAAAAACCTTTTCAACGGAATATATGCG 674

QY 499 GGAGCTATTATGCGGCGAATGTACTATAGTGGCACTACGATTCGCTCTTCTAT 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 GGTGCTATCCACGCAAAAACCTCTTCAATTACAGGAGCTACAAATGTGAGCTGTCTTCT 734

QY 559 CAGAATGACGCCACTTT---TGGAGGTGCTATTCATCTTCAGGTCGCCCTACAGATTGCA 615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 735 GAAATATACCTCCCAAGAAAGAGCGGACATTCAGACTCCGATCCCTTACCATTAAGT 794

QY 616 GTAAATCAGGACAGATTAAGATTTCGACAAATATCGCCCAAGATGTTCTGAGAGGCGT 675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 795 GGAACCAAGGGAAGTCTCTTTTGTGACAAATATCTTTCG---ATATCTGAGAGTGTCA 851

QY 676 TTGTACTCCGATGCTGATATATGATTTGATGACAGATGCTATGTTCTATTTGAGAAAT 735
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 852 ATTTTACAGAACCTTCGCTGACTATTTCTAATAATAGCTAAGATTTCCTTTATGACAAAT 911

QY 736 GAGC---CATGTACTACTGTATATAGTAAAGGAGGCGCTGTCTGTCTCCACTTCA 792
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 912 AAGGTACAGAGGACGAGCTCCCTACACAGGGGAGATATGTCAGAGGTGCTATCTGTGCT 971

QY 793 GGAAGTAGTACTCCACTTCTATTTGACTTCTCTGACAAATTAACAGATTAGCTTTGAA 852
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 972 TATAAACTAGTACAGATTAAGTACACCTTCACCGTACGGAATCAGATGTTACTCTTCAG 1031

QY 853 AGAACCATTCATTAATGGTGGCGGAGCATTTATGTAAGAAACTATGACATCTCTTCA 912
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1032 AACAATACATGCACACAGCGGAGGAGTATCTATGTGAAAAAGCTCGAATGCTCTCC 1091
```

OY	913	GGAGCTCCTACTCTAATTATATCAATATATATCAATATGCAAAATTCGCCAAATTTTGGTGGG	972
Db	1092	GGAGAGCTTACCCTTAATTCATATAAAATATGTCATAGAGGAGTACAGCTCTTAAGATGGGA	1151
OY	973	GCATTTCCCAATGTATCTCGAGAGGGAGATCAGTTTATCGACGAGAAAGAAACATTTACA	1032
Db	1152	GCATTAACCTATCGAAGATATGGGGAAATGAGTTTATCCGGCATATGTTGTGATCTTGTG	1211
OY	1033	TTCCAGAGAAACGGGACGAGCTTAC-----CGTTTTGAAATGGCATTCACCTT	1080
Db	1212	TTTTTAGGGAATACAGCATCTCTACTACTCCGGGAGCAATAGAAATGATATGACACTTA	1271
OY	1081	TTTACAAATCTTAATTTCTCGAAATTTACAGCGGAGAAATGGATGCTCTATAGAAATTTAT	1140
Db	1272	GGAAACGAGTCCAAAGATGAGACGTTTTGCGTTCTGCTGCTGAGAGCCATCTACTTCTAT	1331
OY	1141	GATCCATTTACTTCTAGACAGATGGGTCCTTACCATTGAAATATACAGGAGATCCATAA	1200
Db	1332	GATCCCAATTAACCTACAGATCATCCACACAGTTTACAGATGTCCTTTAAAGTTAATGACACT	139D
OY	1201	AATTAAG-----AGTACACAGGAGACCATCTCTTTTTCGGAGAA-----	1240
Db	1392	CCGGCACAATTCGTGCATACAAATTTACAGGAAACATCATCTTACAGAGAAAGTTATCA	145L
OY	1241	-AGAGCTACCAACGATCCTAGGATTTTAAATCTACAAATCCCTACAGACGTCACCTG	1299
Db	1452	GAGACACAGGCCCGAGATTCTTAAAAATCTTACTCGAAGCTACTACAGCTGTAACTCTT	1511
OY	1300	TCCTCAGGATATCTTAGTTATTTAAAGGGGGGCCGAAGTCACAGTTTCAAAATTCAGGAG	1359
Db	1512	TCAGAGAGTACTGTATCTTTTAAACATGAGAGTACTCTGCACAGCTCAGGACATCTCA	1571
OY	1360	TCCTCAGAGATGCAATTTAGTTTAAAGTTTAAAGAACCAACTGATGACCTTAAGGAAGAC	1419
Db	1572	CAGGACAATTTCTGCTCTCGAAATGAGACGTATGAGACTACTCTGAACCTGCT-----GAT	1625
OY	1420	ATTGCGATCACAGGCGCTCGAGATATGATATGATATGATAGCTTAAGCTCATCTCAACAGACT	1479
Db	1626	ACTGAGCCACCTAAACATTTTGGTCTATTAACTACATCAGTTCATAGAGGGTCAAGAGAGCA	1685
OY	1480	GTTATTTAACAACAACCGCAATTAACAGATATCCGTGACGCACTGTATGAACTATTC	1539
Db	1686	AAATATGAAACCAAGCTACGTCCAAAAATCTGACTTTATCTGGAACATCACTTATATG	1745
OY	1540	TCGCGCTACTGGCAATGCTATGAGATCTCAGAAATGAGAAATTCACAGCTCCCTTG	1599
Db	1746	GACCGCAGGGGACGCTTATATGAATCATATGTTAAGAAATCCCATGCTCAGACATC	1805
OY	1600	CTCTCTTT---AGAGCTGAGCGCGGGGTAGTGTGACTGTATGCTGAGATTTCTCTA	1656
Db	1806	TTTAGGCTCAAACTTCTGTGAACTGTATACACACCGCAGTACATCCAGATCTCTATATG	1865
OY	1657	CCGGTAGTCCCATATATGTTTTTCAAGCAATGG---AAATTAGCTTGGACAGAACT	1713
Db	1866	GGTAGAATTCATTAACGGCTATTCAGGGAACCTTGGGGCCCAATTTGTTGGGGACAGG	1925
OY	1714	GGAAACAAGTTGAGAAATTTCTCTGGATTTAAATTAATTTATACCTAGACCTGA AAAA	1733
Db	1926	GCTTCTACGACTGCACCTTCAACTGACTTAAACCTGGCTATATTTCTTAATCCGAGCT	1965
OY	1774	GAAAGAAATTTAGTCTCTAATATCTTTGTGGGGGAATGCTGTAAATGTGCAGATCTCTAATG	1833
Db	1986	ATTCGGCTATAGCTCTAATTAAGTATGTAATGCACTTATATGATATTAAGCTCTCTCAT	2045
OY	1834	CAGTGTCAAGAGAACCATGATGAGACTTATACAGACGATTCGAGGGCTGTGATTCGATGGA	1893
Db	2046	TATCTTTATGAGACTGCAAAACGAGGGTTGCGAGGGAGACCTGTCTTTTGTGTGCTGGA	2105
OY	1894	ATTGGGAATTTCTTCCATGTATCTGCTCCGCAAGACAATATATATGCTACGCTATACAGC	1953
Db	2106	TTATCTTAACCTTCTCCATATAGGATATGACAAAAACGACGGGGGTTCGCCATTTTGAT	2165
OY	1954	GGTGGATATGTTTATCTGTAAATTAATGAGATCACACTTAAGCACTATACTTCGATGGCA	2013

Db	2166	GGCGGTTATGTCAATAGAGAAACCTACACTACTGTTCAGATTAAGATTCTTAGTGTGCA	2225
Qy	2014	TTTTCCCAACTCTTTTAGTAGACAGACATAGCGGTTTCCACACGAATACAGATG	2073
Db	2226	TTTTGTGACGCTTTGGAAAGAGATAGAGACTACTTTGTAGCTAAAGATCAAGGTACAGTC	2285
Qy	2074	TATTTAGATCGATCTCTATCAATATACAACTCCCTTAGGAATATTTCCGTATGCT	2133
Db	2286	TACGAGAACTCTCTATTTACCGACCAACGAAC--	2325
Qy	2134	TCGGGTAAACCTATGTAAACGTGGGATTTCTCTAAGAGGTTTCTTCAAAATCCTTT	2193
Db	2336	TTCTCTTCTTCCTGCAACATACAGGCTTTGTCGTGTTATTTCTTACAGAGATTCCTG	2385
Qy	2194	ATGATTTTTCATTTTTTGTGCTTATGTGCATGCCACCAATGATATGAAACAGACTAC	2253
Db	2386	TTTCCTTTTCAGGAACCTTAGC-TACACCATTACGATTAACGATCTGAAACCAAGAT	2444
Qy	2254	GCAATTTCCCTATGTGTGAAACACAGCTGAGAAACATTTGGGGCTATAGATGCGGA	2313
Db	2445	ACAAATATTCCTACTGTATAAGAGACTGGGGAAATGATAGTTTGGCTTATGAAATTCGT	2504
Qy	2314	GGAGCATGCCCTCATTTGTTATTTGAAACGGAAGACTTTCCAAAGGTGCATCCACTT	2373
Db	2505	GGAAGAGCTCCGATTTCTTAGATGAAGCTGTATTTGACAGCTA--CATGGCCCTTC	2561
Qy	2374	ATGAACATACATAGTTTATGTTTATGTTTCCAGGAGATTTCAAAAGACACAGTCAGATGGC	2433
Db	2562	ATGAAATTTGAGTTTGTCTATGACATCTACAGGAAGTTTAAAGAACAGGAACAGAACT	2621
Qy	2434	CGTATGTTTATGATGGAGATTTACATGATTTCTGTACCTCTAGGATACGCTTTGAG	2493
Db	2622	CGTAAATTTGGAAGTACCGCTCTGTGAATCTTGCTTACCTATCGGATCCGATTTGAT	2681
Qy	2494	AAGCTGCATTTCTCAGATGTACTCTATGACTTATGTTTCTCTATATTCCTGATATT	2553
Db	2682	AAGAAATCAGACTCCCAAGATGCAAGTACAACTAACTCTGTGTTATACCTGTGATCTT	2741
Qy	2554	TTCCGTAGAGATCCCTCATGTGAAGCGCTGTGTATAGGGGAGACCTCGGGCTGTT	2613
Db	2742	GTTCGTGTATACCCCGACTATACGACAAACCTCGGATTAGGGTGTATCTTGGAAAC	2801
Qy	2614	CCGCGACAGACAGTATACACATAGCTTTTGTAGGAGTGGAAACGGGTGCGATCACTT	2673
Db	2802	TTCCGTACGATTTGGCAACAAAGCTTTTAGTCTCTCGACAGGGAAACATTTTGGCTT	2861
Qy	2674	AAGCACTATACGTAGCTCTTATGTGAGAGAGTATAGAAATGCCGCCCACTGATAGACT	2733
Db	2862	AACTCAAATTTTGAAGCTTTTACCAATTTTCTTTTGAATTCGTGGCTCATCTGCAAT	2921
Qy	2734	TATTAATTAACCTGTGGAACCAATTTGCTTTTGAAGCTTTCATATG	2782
Db	2922	TACATCTTAGACTTAGAGCAAAATATCCAAATTTCTAATGCGTAGCTTGG	2970

FEATURES	source
LOCUS	2787 bp DNA
DEFINITION	Sequence 9 from Patent WO958953.
ACCESSION	AB1835
VERSION	AB1835.1
KEYWORDS	GI:6731868
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 2787)
AUTHORS	Madsen, A. and Birkelund, S.
TITLE	NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL	Patent: WO 958953-A 9 30-DEC-1998;
	MADSEN ANNA SORIE (DK); BIRKELUND SVEND (DK)
	Location/Qualifiers
	1..2787

/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 811 a 583 c 598 g 795 t  
ORIGIN

Query Match 8.88; Score 270; DB 9; Length 2787;  
Best Local Similarity 48.4%; Pred. No. 1.1e-54;  
Matches 1266; Conservative 0; Mismatches 1260; Indels 90; Gaps 15;

QY 207 CTCTGATTAATAACGCTGCTCAATGTGACTCAGAAAGTGTACGCTTCCGAGAAA 266  
DB 207 CACAGCAATACAAAAGCTGTTTAAACAACCTAAGAGGCGATTGCTTCAAGGTAA 266  
QY 267 TCATCATGGGTATATTTTATATATATTTCTCTGAGAACTACAAAGAGAGGGCTGACT 326  
DB 267 CGGGAAGCTCTATTTGTTCCAAAGGTGATGAGGAGCTGTAT---CGAGGGCTGCTGT 323  
QY 327 TTGTTGCCAAGATCTCAGCAACGCGATTTTCTGGGCTTCCAGCGCTCTCTTTAT 386  
DB 324 TAACAGCAGGGTGTAGTAAATCTACAGGTTTATAGGTTTCTTCCGCTATCTTTAT 383  
QY 387 TCAGAGCCCGGAGATTTAA-----GAAAGAGATGCTCTATTCAAAATATGC 437  
DB 384 TCGCTCTCTGAGAGTTCGTAACCTACCGCAAGAGCGCTTATGCTCTACAGGGTAG 443  
QY 438 ACTTATGCTCTTAAACAATTATGATGCGTTTGAACAAACCAAGTAAGACTTAAG 497  
DB 444 CTGAAGTTTGACAAAATGTCAAGTTTCTCTTACGCAAAAACCTTTCAAGGATATAG 503  
QY 498 CGGAGCTATAGTGGGGCAATGTTACTATAGTACGCACTAGATTCGCTCTTTCTA 557  
DB 504 CGGTGCTATCAGCAAAAACCTTTCTTACAGAGGAGTACAAATGTCACTCTGTTTC 563  
QY 558 TCAGAAATGCAGCACTTT---GGAGGTGCTATCCATTTCTCAGTCCCTTACAGATTC 614  
DB 564 TGAATATACCTCTCAAGAAAGAGCGGAGCCATTCACACTCCGATCCCTTACATTAAC 623  
QY 615 AGTAATCAGGCAAGATTAAGATTGCAAAAATCTGCCAAGATGTTCTGAGAGGGC 674  
DB 624 TGGAAACCAAGGGGAAAGTCTTTTCTGACAAATCTTCTGCG---GATTCGAGTGC 680  
QY 675 TTTGTACTCGATGTGATTTGATATATGATTCAGAAATGCTTATGTTCTATTCGAGAAA 734  
DB 681 AATTTTACAGAAAGCTCGGTGACTATTTCTAATTAAGCTAAAGTTTCTTTATGCAAA 740  
QY 735 TGAG---CATTCGCTCTCTATAGTAGGAGGGGCTGTCTGTTGCTCTCCCAATTC 791  
DB 741 TAAGGTACAGGAGCGAGCTCTCAACACAGGGGATATGTCAAGAGGTCTATCTGTGC 800  
QY 792 AGGAAGTACTCTCAAGTCTTATGACTTCTCTGACAAATAAACAGTTAGCTTTGA 851  
DB 801 TTATTAATAAGTACAGATTAAGTACCTACCTGACCTGGAATACAGATTAAGTCTTCA 860  
QY 852 AAGAAACATTCATATATGGGTGGCGGAGCCATTTATGCTAGAACTTATGATCTTTC 911  
DB 861 CAACATATATCACAACACCGGAGAGCTATCTATGTAAAAAGCTCAACATGTGCTTC 920  
QY 912 AGGAGTCTCTATTTTATCAATATATATATATATGCAAAATTCGCAAAATTTAGGTG 971  
DB 921 CGGAGAGCTTACCTTATTCAGTAAATATGATGCAATGAGGTACAGCTCTTAAAGGTG 980  
QY 972 AGCTATGCTATGATAGTGGAGGAGATCACTTATCAGCAGAAAGAAACAAATTAAC 1031  
DB 981 AGCATAGTATGAGATAGTGGGAATGAGTTTATCGCGGATAGTGTACATGTT 1040  
QY 1032 ATTCAGAGAAA-----CCGAGAGACTTACGCTTTTGTGATGGCATCTCT 1079  
DB 1041 CTTTTTAAAGTAACAGTCACTTCTACTACTCTCTGAGCAATATAGATGTATCGCTT 1100  
QY 1080 TTACAAAATGCTAAATTCGAAATTAACAGGCGAGAAATGATGCTATTAATTTTA 1139  
DB 1101 AGGACGAGTGCAGAAATGACAGCTTGTGCTGCTGTAGAGCCATCTACTTCTA 1160

QY 1140 TGATCTATTACTTCTGAG-----CAGATGGGTCTACCAATTTGAATATCAACGAGA 1193  
DB 1161 TGATCCATATACAGAGATCTTCCACACAGTATAGATGTCTTAAAGTTATGAGAC 1220  
QY 1194 TCTTAAAAAT-----AAGATACACAGGAGCATCTCTTTCTTGAGAAA----- 1240  
DB 1221 TCCGGCAGATTTCTGACATACAAATATACAGGAAATCATCTTCTACAGAGAAATGATC 1280  
QY 1241 --AGAGTGTGCAAGAGATCTAGGGATTTTAAATCTACAAATCCCTGAGAAAGCACT 1298  
DB 1281 AGAGCAGAGCGCGACATTTCTAAATCTTACTTTCAGACTCTACAGCTGTAACTCT 1340  
QY 1299 GTCTGAGATTTACTTATTTTAAAGAGGGGCGGAGATGACAGTTTCAAAATTTACAGCA 1358  
DB 1341 TTGAGAGAGTACTTACTTAAACATGAGATGACTCTGACAGCTCAGGCAATTCATCTCA 1400  
QY 1359 GTCTCCAGATTCGATTTAGTTTATGATTTAGAACCAACCTGATAGCTCTTAAGGAGA 1418  
DB 1401 ACAGGCGATTTCTGCTCGAAATGAGCTAGGAACTACTGTGAACCTGCT-----GA 1454  
QY 1419 CATTCGATCAGAGGCTCGGATAGATATAGATAGCTTAAGCTATCCCAACAGCAGC 1478  
DB 1455 TACTAGCACAATTAACATTTTGTCTATTAATCAGTCTTATAGAGGTGCAAGAAAGC 1514  
QY 1479 TGTATTTAAAGCAAAACCGCAAAATTAACAGATATCCGTAGCGACTTATAGAACTTAT 1538  
DB 1515 AAAAATGAAACCAAAACCTACGTCAAAAATCTGACTTTATCTGGAACATCACTTATTT 1574  
QY 1539 CTGCGCTACTGCAATGCCATTAAGATCTCAGAAATGAGAAATTCAGAGCTTCCCTCT 1598  
DB 1575 GAACTCAGCGGACGTTTATTAATCATGTTTAAAGAAATCTCAGTCTACAGCAT 1634  
QY 1599 GCTCTCTTT---AGAGCTGAGCGCGGGGTAGTGTGACTGTATGCTGTGAGATTTCT 1655  
DB 1635 CTTAAGACTCAAAAGCTCTGGAAGTCTTAACAGAGCAGCTGATCTCAATCTTAAT 1694  
QY 1656 ACCGATAGTCCCATTTATGTTTCAAGCAATGG---AAATTAAGTTGACAGAGAAC 1712  
DB 1695 GGGTGAAGAAATTCATTCAGCTATCAGGGAATCTGGGGCCCAATGTTTGGGGGAGC 1754  
QY 1713 TGGAAACAAGTTGGGAATTTCTTGGGATTAATTAATTAAGCTTACGCTGAGAAA 1772  
DB 1755 GCGTTCTACAGCTGCAACCTTCAACTGAGCTAAACGCTATATTTCTTAATCCGAGCG 1814  
QY 1773 AGAAGAAATTTAGTCTTATATCTTGTGGGGAATGCTGTAATGTGATCTTAAT 1832  
DB 1815 TATCGCTCTTATGCTCTTATAGCTTATAGCTTATAGATATATAGTCTCTCCA 1874  
QY 1833 GCAAGTTCAAGAGACCCATGATCGAGCTTACAGACAGATCGAGGCTGTGATCGATGG 1892  
DB 1875 TTATCTATGAGACAGCAAGAGGTTGAGGAGACCGTCTTTTGGGTGCTGG 1934  
QY 1893 AATTTGGAATTTCTTCAATGATCTGCTCCGAAGCAATATATAGTACCGTATTAACG 1952  
DB 1935 ATTTATCTTACTTCTTCAATAGAGATGATCAAAAACAGAGCGGGTCTTCCCATTTGAG 1994  
QY 1953 CGGTGATATGTTCTATCTTAATATATGATCAACCTTAACCATTAATCTGATGGC 2012  
DB 1995 TGGCGGTATGATAGAGGAAACCTACTACTTGTTCAGATTAAGATTTTATGTGCTGC 2054  
QY 2013 ATTTTCCCAACTCTTTAGTAGAGACAGACTATGCGGTTTCCAAACGATACAGAT 2072  
DB 2055 ATTTTGCACCTCTTTGAGAGATAGAGACTTGTGTGTAAGAAATCAAGGTACAGT 2114  
QY 2073 GATTTAGATGCTATCTCATATATACAACTCCCTTAGGAAATTTTCCGTTATGC 2132  
DB 2115 CTACGAGGAACTCTTATTAACAGCAACAGAAAC-----TAT 2154  
QY 2133 TTGCGCTAACCTTAATGTAAGCTGAGATTTCTCAAGAGGTTTCTTCAAAATCTCT 2192  
DB 2155 ATCTCTCTTCTTGCAAACTACGCGCTTGTGTGCTTATGTCTTACAGAGATTCCT 2214





Db 2792 GACGATGACGATACACACGCTGCTTTTACGACATGCTGTTCCGGTTCATCTGA 2851  
 QY 756 AGGTAAAGGAGGGGCTGTCTGTCTCTCCACTTCAGAGTAGTACTCCAGTCTCAT 815  
 Db 2852 TGGTTGCGGTGGAGCTATCCATGCT-----AGCAAAACAGGCTTCAGCAGCAG 2899  
 QY 816 TGTGACTTTCTTGACAAATTAACAGTGTCTTTGAAAGAAACCATTCATTAATGGGTG 875  
 Db 2900 CCTTACTATAGAGATTAACAAAGCTGTGATTTTGAAGAAATACTCTTCAGCAAAAG 2959  
 QY 876 CGGAGCCATTATGCTAGAAACTTACATCTCTTCAGAGGTCTTACTCTATTTATCAA 935  
 Db 2960 TGGAGGATTTTACACCGATTAACCTATTTGACTCTGCGGGCCCTACGCGCATTTATCAA 3019  
 QY 936 TAATATATCATATGCAAAATTCGCAAAATTTAGTGGAGCTATGCTCCATGATCTGGAG 995  
 Db 3020 TAACAAAGT---TACCATCTACACCTAAGGGGAGCTATGTGGATATCTGCAATGG 3076  
 QY 996 GGAGATCAGTTTATCAGACAGAGAAAGAACATTCATTCACAGAAACCGAGAGCTT 1055  
 Db 3077 AGAATGTAGCTTACCGCTACACATGGGATATTTACTTTGATTAATACCTGATGGCCAC 3136  
 QY 1056 AC-----CGTTTTGAATGGCATCCATCTTTTACAAAATGCTAAATCTCT 1100  
 Db 3137 ACAGACAAATGCTACAAATAAAAGAAATGCCATTAACATTTGAAGGCAATGTAATTCGT 3196  
 QY 1101 GAATATACAGGCGAGAAATGGATGCTCTATAGATTTTATGATCCATTTACTCTCTAGAC 1160  
 Db 3197 CAATCTACGTCGACGCTGTGAAAGACGATTTCTTCTATGATCTCTACACAGTTGAGG 3256  
 QY 1161 AGATGGGTCTACCCAAATGGAATATCAACGAGATCCCTAAATTAAGATACACAGGAC 1220  
 Db 3257 TAATGCTGCTGATCTCTCTCTGATTAAGCTGAGGGTATTAACGATATATAGGAG 3316  
 QY 1221 CATACTCTTTCTGAGAAAGAGCTCTAGCAACGATCCTAGGGAT-----TTAA 1271  
 Db 3317 AATTAATTTTCTCAGGAAAGAAAGCTCACTGAAGAACAACTGCTGTTCCGATTAACCTAAA 3376  
 QY 1272 ATCTACATCCCTCAGACGCTCAACCTGTGAGAGATTTAGTTATTTAAAGGGGGG 1331  
 Db 3377 GACACACTTTACACAGCCTATCATTAGCTGCTGGTGAAGCTGTGTACGACGGGTGT 3436  
 QY 1332 CGAATCACAGTTTCAAAATTTCAAGCAGTCTCCAGATCCATTTAGTTTAAATTTAG 1391  
 Db 3437 GGAAGTGAAGCAAAACAGTCCGTCGCAACAGAGATCTTTGATTCGATGATGACG 3496  
 QY 1392 AACCAACTATATGCTCTTAAGGAAAGATTCGATACAGGCTCGCGATAGATATAGA 1451  
 Db 3497 CACAAGTTATCCGCAAAAACAGAAAGTGTACAGTACGATCTGGCTATTTAATCCGAA 3556  
 QY 1452 TAGCTTAAGCTCATCTCAACAGCAGCTGTTATTAAGCAAAACCGCAATTAACAGAT 1511  
 Db 3557 TACTTA---GATGGGAAAAATTCGCCGATGTCGATGCCGTGTGCTGGGAAAGATGT 3613  
 QY 1512 ATCCGTGACGAGCTATATAGAACTTATCTGCTACTGCAATGCTTATGAAAGATCTCAG 1571  
 Db 3614 GACTTATATCAGGCTATGTCGCTTATGATCCTACAGGAAAGTTTATGAAAAACCTATA 3673  
 QY 1572 AATGAGAAATTCACAGACGTTCCCTC---TGTCTCTTTAGAGCTTGAGCCGGGGTAG 1628  
 Db 3674 GCTAAATGATAGCTTGAAGCTTTAGAGAAATTCACATTTCTGGGAAAGGTTCCGGTACAC 3733  
 QY 1629 TGTAGCTGTAACGCTGAGGATTTCTACCGGTAAGTCCCATTAATGTTTCAAGGCA 1688  
 Db 3734 AACCAAGCTGCTGATGTCATGTGTGTGTGCTGAAACCCCATATGTGTTATACAGAAA 3793  
 QY 1689 TTGAAATTAGCTTG-----ACAGGAATCGAAACAAAGTGTGAGATTT 1733  
 Db 3794 CTGCTCTGTAGTTGGGTCAAGATATATACTGTGATCCTAAAAACAAACAGCAATCTT 3853  
 QY 1734 CTTCGTGGGATTAATTAATTAATAGCTTACAGCTGAAAGAAAGAAATTTAGTCTTAA 1793  
 Db 3854 TACCTGGAATTAACAGGATATGTTCCAAATCCTGAACGTGTGCTCCGCTAGTACTCAA 3913

QY 1794 TATCTTGGGGGAAATGCTGTAATGTCAGATCCTTAATGCAAGTTCAGAGACCCATGC 1853  
 Db 3914 TAGCTTTGGGGATCCCTTATATGATTTAGCTTCTATTAAGATGTCTTGAAGAGTAGT 3973  
 QY 1854 ATGAGCTTAC---AGACAGATCGAGGGCTGTGGATCGATGATTTGGGAAATTTCTTCA 1910  
 Db 3974 TGATGATATCTTGAGACAGCTGTGGTTTGTGGTCTCTGAAATGGGAACTTCTTCA 4033  
 QY 1911 TGTATCTGCTCCGAGACAAATATAGTACCTCATTAACAGGGGTGAGATATGCTATAC 1970  
 Db 4034 TAAAGATCGGAATGCTGAAATGCAAAATTCGTCATATGATGCGGATGTGGAGG 4093  
 QY 1971 TGTAAATATGAGATCACACCTTAAGCCTATAGCTGATGATTTTCCCACTCTTTAG 2030  
 Db 4094 AGCCACAAACAAATACCTCGAGAGAGATTTCTCTTAGTGTGCTTGTGAGTATTTG 4153  
 QY 2031 TAGAGACAGAGCTATGCGGTTTCCAAACAGAAATACAGATGATTTAGATGCTATCT 2090  
 Db 4154 AAAAGATTAAGACTACTTGTAAAGCAAGACGCCCAACGCTATATCGGGTCTGTATA 4213  
 QY 2091 CTATCAATATCAACCTCCCTAGGAAATATTTCCGTTATGCTTCCGATACCTTATGT 2150  
 Db 4214 TTATCAGCAT-----GTGAGCAAGTTGATGATCTCAGCGGTATTTAATGG 4261  
 QY 2151 AAAGTGGGATTCCTCAAGAAAGTTTCTTCAAAATCCCTTATGATTTTTCATTTT 2210  
 Db 4262 GCCTACACGCTGTGTTGCT---AGGTTTCTTAAAGATTCCTATTTCTGTGATGACA 4318  
 QY 2211 GTGTCTTATGTCATGCCAACCAATGATATGAACACAGCTACGCAATTTCCCTATGT 2270  
 Db 4319 AATTAACCTATGTCACAGGCGCAACACATGACACACTCTTAACAGACTCTGTAAGT 4378  
 QY 2271 GAAAAAGCTGAGAAACAAATGTTGGCTATAGAGTGGGAGGAGCATGCTCAT 2330  
 Db 4379 GAAAGGTTCTTGGGGTATATACCTGGGCTTAACCTTCTACTATAGCTACTATCC 4438  
 QY 2331 GGTATTTGAGAAAGGAAAGACTTTTCCAAAGTGCATCCATTTATGAAACATAATAGT 2390  
 Db 4439 GGTATTT---AGTTCTCTATCTTGTGATGATTTATGACCGCTTGCAAAATTAACAAAGTGT 4495  
 QY 2391 TTTATCTTATCAGGAGGATTTCAAGAGACGACTCGAGATGCCGTAGATTTAGTATAG 2450  
 Db 4496 CTATGCGCACCAAGATGACTTTAAAGAACCAACACAGAAAGCGCGGTCTTTGAAACAG 4555  
 QY 2451 GAGTTTAACATGCAATTTCTGACCTGATAGGCAATACGCTTGAAGAGCTGACATTTCTCA 2510  
 Db 4556 CGATCTTCTAAGCTTCTGTACTTACCTATAGTATTAATTTGAGAACTCTCTATGAGA 4615  
 QY 2511 GGATGATCTATGACTTTAGTTTCTCTATATTTCTGATATTTTCCGTAAAGATCCCTC 2570  
 Db 4616 GAGAAAGCTTATGATCTTACAGATGATATATACCTGATGTATACCTCATATCCAG 4675  
 QY 2571 ATGTGAAGCTGCTGTGTGATTAAGCGAGACTCTGCTGTGTTCCGACACACAGATAC 2630  
 Db 4676 CTGTATGACAGGATTTGGCGATGATGACATGCTGCTGTATTAACACACACTGAGATCTTG 4735  
 QY 2631 AAGCATGCTTTTGTAGAGGTGGAACGGGTCGTATCACTTTAAGCATATACTGAGCT 2690  
 Db 4736 TAGACAGCTTTTATGATTTGCGCGGGTAAACATATTTGCTTAACCTCTGTGTGAGAT 4795  
 QY 2751 AAGCAATTTCTTTTGAAGGTTTCC 2778  
 Db 4856 CGCTAAGCTGCGTTCTTAAAGCTCC 4883

RESULT 11  
 AE002193/c  
 LOCUS AE002193 10026 bp DNA BCT 30-MAY-2000



DEFINITION	Chlamydia pneumoniae AR39, section 26 of 94 of the complete genome.
ACCESSION	AE002193
VERSION	AE002161
KEYWORDS	AE002193.2 GI:8163406
SOURCE	
ORGANISM	Chlamydia pneumoniae AR39.
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
AUTHORS	1 (bases 1 to 10026) Read,T.D., Brumham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE	Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
JOURNAL	Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE	20150255
PUBMED	10684935
REFERENCE	2 (bases 1 to 10026) Read,T.D., Brumham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
AUTHORS	Direct Submission Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA On Jun 1, 2000 this sequence version replaced gi:7189234.
COMMENT	location/Qualifiers
FEATURES	1..10026
SOURCE	/organism="Chlamydia pneumoniae AR39" /strain="AR39" /db_xref="taxon:115711" /note="synonym: Chlamydia pneumoniae AR39" complement(105..2897) /gene="CP0307" complement(105..2897) /gene="CP0307" /note="similar to GP:4376729, identified by sequence similarity; putative" /codon_start=1 /transl_table=1 /product="polymorphic membrane protein G family" /protein_id="AAE38164.1" /db_xref="GI:7189235" /translation="MKRIKRLIISSTLVPIILSTIATYGADASISPTDSFGGSGSTFPRKSTADANGTVVLSGNYINDAGKGLTGGCFPTGGDLTFPKGYSFSPNVDAGSAGAASTADAKALFTFGFSLITADAGTVASGKSTLSAGALNLINDNTIFLSONVSEANNNGGATITAKTLISGNTSITFTSGNSAKRLGAIYSAASISGNTGDLVPMNKGETGGAGLGEFASSITONSILFSGNTPADGAGCAIYCKEIGETPTLITISGKSLTFEANSVTCGACILCAHGLDLSAAGPTENNRGCTAGKGAAYIADSGSLSISANGDITPLIGNTLISITSAPTSTNAIYIGSSAKTINRAAGOSIYIYDASINITYASDYLITINPDNSPDIYSGTIVSEKLSADEKADNFTSITIKOPALASGITLALGNELDVNGFTGTEGSTLMDPGKILADVAEILSLKIVLDSLEGGKSVSIEFGAGNKITLTLSPLFVODSSGNFEESHINAFDPLVAFATASDIYDALITSPVLOTEPMHYGOGHEATWADSTIAKSGMTWTTGYNPNPERASVDPDSLMASTDIRRQOEDINYSOANSIYIOGRLWASGTANFPHDKSTNOAFRHSKYIVGSGADEFSENEITFVARFOLFGKDKDLFIVENTSHNYTSLXLYGHRAPLGLPMPGSDITMDKIDIPILIAQLSITTKKNDMDRTIYSTYPEKSSWTNNSGALGLALYLRKAPFFDTGIPFFFLFQVAYSIRKQNFESGAGARAFDDGDLVNCSTIPVIRLEKISIDERNNEISLAYTGDVYRKNRPSRTSLWVSGASWTSLCKNLARQAFILASGSHLITSPHYELSGEAAVEIDRSGSAHYIVVDCGLIRSF"
gene	complement(2941..5751) /gene="CP0308" complement(2941..5751) /gene="CP0308" /note="similar to GP:4376728, identified by sequence similarity; putative" /codon_start=1 /transl_table=1 /product="polymorphic membrane protein G family" /protein_id="AAE38165.1"
CDS	

Query Match	Best Local Similarity	48.8%	Score	260.8	DB 1:	Length	10026
Matches	1055	Conservative	0	Mismatches	1033	Indels	75
				Gaps			10
QY	657	GAAGTTCCTGGAGGGCGTTGTACTCCGATCGATGATATGTGATATGTGATCAGAAATGCTTA	716	DB	5074	GAAAGGATCTGGCGGTGCTGTACAGCAGTACAACTGTTGATACAGATATCAATCTTCA	5015
QY	717	TGTTTATTTTCGAGAAATGAGGCAATTGACATCTCTATAGTAAAGGAGGGCGCTGCTG	776	DB	5014	AGTGATCTTTGACGCAATAGTGTGGGAAAGCCCTCAACTCAGGGCGGCGCTATTATG	4955
QY	777	TGTCTTCCCACTTGAGAAAGTAGTACCTCACTTGTGACTTCTTCGACAAATTA	836	DB	4954	TTGC-----ACACGACAGATTAACAGATGACTCTTACTGGGACAA	4913
QY	837	ACAGTTAGCTTTGAAAGAACCATTCATATAGGTGGGCGAGCCATTTATGCTAGAA	896				

D	4912	AAACCTCTCTTCCAAATAATATACGATTTGACATATGCGAGGACCATCTCGACCTCA	4855
O	897	ACTTAGCATCTCTTCAGAGAGTCTCTACTTATTATCAATAATATATCATATGCAAAATTC	956
D	4852	GSTCGATATTTCCCTGGAGGTCTCTACTCTATTATCAAGATATATCTCGAAGTAGCGC	4793
O	957	---GCAAAATTTAGSTGAGCTATTGGCATGTGACTGGAGGGAGATAGTTATGAC	1013
D	4792	CGGTACGGGAGGAGGAGAGGAGCGATCAATATAGCATCTGCTGGGAACTGCTCTCTGCG	4733
O	1014	AGAGAAAGAACTTACATT-----CCAAGAAACCGGACGAGCTTACCGTTTT	1064
D	4732	TACTCTGTGAGATATTTACTTCAATAACCAACGATCCACCAAGGAAAGCAATGACAA	4673
O	1065	GAATCGCATCATCTTTTACAAAATGTAATTCCTGAAATTTACAGCGCAAGAAATGATG	1124
D	4672	AAAGCAATTAATATCATATGATACCGCTTAAGTCATCGATCGAGCTGCTACGGGGCA	4613
O	1125	CTCTATGAATTTATATATCTTATATCTTCTTGAGAGATGGGTTCACCC-----	1174
D	4612	ATCTATCTATTTCTATATATCCATCAACAAATCCAGGACCGAGCTTCTACGCAACATT	4553
O	1175	--AATGATATTCACAGGAGATCTTAATAATTAAGATACACAGGACCATCTTTTC	1232
D	4552	GAACCTAAACTTACACATCCGAACGTGAGTCGATGCGGGGTGCGATGTCTTTTC	4493
O	1233	TGGAGAAAGAGTGTACCAACGA-----TCTTAGGATTTTAATCTACAACTCC	1283
D	4492	TGGAGAAAGAGTTTCCCTCTACAGAAAAAGCAATCGCTGAAACGTACCTCTACATATCG	4433
O	1284	TCAGACGTCAACCTGTCTCAGGATATCTTATTATTAAGAGGGGCCGAAATCACAGT	1343
D	4432	ACAACCTGCAGTATTACCGGGGAGATCTTGACTCTGATGAGATCACCGCTAACCTT	4373
O	1344	TTCAAAATTCACGAGCTCCAGAGTCCAGATCGCTTATAGTTTAGTTAGAACCAACTAT	1403
D	4372	CAAGATCTGACTCAAGTCTCAGAGATCCCGCATCTTAATGATGGGGGACACTACCTAG	4313
O	1404	AGCCTTAGGAGAGACATTGCCATCACAGGCTCGCGATAGATATAGATAGCTTAAGTCC	1463
D	4312	TGCTAAAGAGCAATCTTTCGCTTATAGCTTAGACGTAATAATCTCTCTTAGTAGG	4253
O	1464	ATTCCTCAACAGCAGCTCTTATTTAAGCAAAACCCGAATTAACAGATATCCGTACCGA	1523
D	4252	AAACCAACAGGACACTTT---AAAAACGAGAGTGCAGATTAATAATATCAGCTATGCGG	4196
O	1524	CTCTATGAAATTTCTCGGCTACTGGCAATGCTTGAAGATCTCAGAAATGAGAAATTC	1583
D	4195	AACGATGGCGCTTATTCACAGGAAAGGTCATTTATGATGATCATTACTTAATAAGTGC	4136
O	1584	ACAGACGTTCCCTCTGCT-----CTCTTAGAGCTGGAGCCGGGGTAGTGT	1631
D	4135	TAGTACTATCTCTCTTCTTGAACCTACCAACCGAGAGCAACGAAAGATATCTGCG	4076
O	1632	GACTGTAACTGCTGAGAAATTTCTTACCGGTAAGTCCCAATTTGTTTCAAGGCAATTG	1691
D	4075	AGCTCTTTTACCTGACTCTTTAABAACCTGAAACCCACTACGGGTATCAAGAAATCG	4016
O	1692	GAAATTAGCTTGACA---GGACTGAAAAAAAGTTGGAGAAATTTCTTGGGATTAAT	1748
D	4015	GCACTGTCTTTGGGCAAAATCAACATCTCAAAAATAGAAACATCAACTGACCCGTA	3956
O	1749	AAATTAATAGCTGAGACCTGGAATAAAGAAATTTAGTCTCTATATCTTGTGGGGAA	1808
D	3955	AGGATCAATTTCTATGCTCTGAGGAAAAAGATATCTCCCTCAAAATGCTTATGGGAAA	3896
O	1809	TGCTGTAAATGTCAGATCTTAAATGCAAGTTCGAAGGACCCCTGCATGACCTTACAGAC	1868
D	3895	CTTTATATGATATACGCTCATATACAGCTTTATAGAAACCAAGTCCATGGGGACCTTT	3856
O	1869	AGATCGAGGCTGTGATGATGAAATGGAAATTTCTTCATGTATCTGCTCCGAGA	1928

Db	3835	TGAGGAGAGCTATGCGCTTCAGGAATATGCAATTTCTCTATAGAGATTATATGCCAC	3776
Qy	1929	CAATATAGGAGTACCGCTATTAACACGGGTGATATGTTCTATCTGTAATATATGAAATAC	1988
Db	3775	CCGGCATGGTTCCGCATATCAGCGGGGGTTATGACACTAGGAGATACGCAACAATCC	3716
Qy	1989	ACCTAAGCACTATCTTCGATGGCATTTTCCAACTCTTAGTAGAGACAAAGACTATGC	2048
Db	3715	TGCCGAGATACGTTACTTTGCTTCTGCGACGCTCTTTGCTAGAGATCGCAATCAT	3656
Qy	2049	GGTTTCCACAACGATACGAATGTATTTAGATCGATCTCTATCAATATACAACCTC	2108
Db	3655	TACAGGTACAGACCCAGGAGATACTTACGGTGGCTTTGTATTTCCACATACGAAG	3596
Qy	2109	CCTAGGGAATATTTTCGCTATATCGTCGGTAAACCTATATGAACGTCGGATCTCTC	2168
Db	3595	GCTCTTGACATCTCCCAATTTCTCTGGGAAAAGACCCAGCTCCCTGGGTCTCTC	3536
Qy	2169	AAGAAGTTCTTCCAAATATCCTTATATGATTTTTCATTTTTTGTGCTTATATGTCATGC	2228
Db	3535	TGAGATCTCCAGATCATCTCTTATGCG--TTGCATGTCTAAATTCAGTTATCTCATAC	3479
Qy	2229	CACCAATGATATGAAAACAACACTACGCAATTTCCCTATATGTTGAAAAAACAGTGGAGAA	2288
Db	3478	AGACAAACCAATGAAAGCAATATATACCGTAACTATACATCAACAAAGGTTCTTGAGAAA	3419
Qy	2289	CAATTTGGGCTATATAGTGGGAGGAGACATGCTCTATTTGTTATTTAGAGAACGGAAG	2348
Db	3418	CGATGCTCTGTCAGATCTTGAGACTAGCGCTCT--TTGTATTTTCGGTTCGTA	3362
Qy	2349	ACTTTTCCAAAGTGCATCCCATTTATGAACACTACATTAATGTTTATGCTTATCAGGAGA	2408
Db	3361	TCTTCTATAAAGAAATCGAACCTTTGTCTCAAGTACGTATCTATCTATCGGCATCAGACAAGA	3302
Qy	2409	TTTCAAGAAGACGACTCAGATGGCGGTATGTATAGTAATGGAGATTAAACATCGATTTTC	2468
Db	3301	CTTCTAGACGGTTATGCTTAGAGACGGCTTTCATATAAACGAGCTTATCAACGTAGA	3242
Qy	2469	TGTACTCTTAGGCATACGCTTTGAGAAGCTGGCACTTCTCAAGATGTACTCTATGACTT	2528
Db	3241	GATTCCTATAGGGCTCTCCCTCCAAAGAGACTCAAAATTCGAAAAGGGAACTTACGATCT	3182
Qy	2529	TATGTTTCCATATTCCTGATATTTTTCGTTAAGATATCCCATGTAAGCGCTGCTGCT	2588
Db	3181	TACTCTTATGTAATATCTGATCTTACCGAGCGCAATCTTAATATCAACTTCCCTAT	3122
Qy	2589	GATTAGCGAGACCTCTGCGCTTCTCCGACAGACACGATACAGACATGCTTTTGTAGG	2648
Db	3121	AGCTAAGGATGTAACTGATGGCTGATATGTTACAAACCTCGCACGACAAGGTTTTCGT	3062
Qy	2649	GAGTGAACGGGTGCGATACCTTTAAGCACTATCTAGCTCTTATGTCGAGAAAGTAT	2708
Db	3061	TCTGGCTGCACACCATTTCCAACTGAAACCCCAATGTAACCTTCGGCTCAATTCGCTT	3002
Qy	2709	AGATATGCGCCCATCTGATGATATATATATTAATTAACCTGGAACCAATTTGCTTTTA	2768
Db	3001	TGAAGTACGAAGTCTTTCACGAATATATATACAAACCTAGGCTCTAAGTTTGTCTTA	2942
Qy	2769	GA 2770	
Db	2941	GA 2940	
RESULT 12			
AE001627			
LOCUS	AE001627	15068 bp	DNA
DEFINITION	Chlamydia pneumoniae section 43 of 103 of the complete genome.	BC1	01-DEC-2000
ACCESSION	AE001627	AE001363	
VERSION	AE001627.1	GI:4376721	
KEYWORDS	Chlamydia pneumoniae CWL029.		
SOURCE	Chlamydia pneumoniae CWL029		
ORGANISM	Bacteria; Chlamydiales; Chlamydiaeae; Chlamydia.		

RESULT 12

AE001627

LOCUS	AE001627	15068 bp	DNA	BCI	01-DEC-2000
DEFINITION	Chlamydia pneumoniae isolate 43 of 103 of the serotype genome				

DEFINITION	Chlamydia pneumoniae section 43 01 103 01 line complete genome
ACCESSION	AF001637 AF001363

ACCESSION	AE001027	AE001303
VERSION	1	GT:4376721

**KEYWORDS:** adolescents; delinquency; family; intervention; juvenile justice

SOURCE Chlamydia pneumoniae CWL029.

ORGANISM Chlamydia pneumoniae CWL029

Bacteria; Chlamydiales; Chlamydiaceae; Chlamydothrix.

REFERENCE 1 (bases 1 to 15068)  
AUTHORS Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,  
TITLE Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.  
JOURNAL Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
MEDLINE 99206606  
PUBMED 10192388  
REFERENCE 2 (bases 1 to 15068)  
AUTHORS Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,  
TITLE Grimwood,J., Davis,R.W. and Stephens,R.S.  
JOURNAL Direct Submission  
SUBMITTED (01-DEC-1998) Program in Infectious Diseases, University  
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
FEATURES  
source location/Qualifiers  
1.15068  
/organism="Chlamydia pneumoniae CWL029"  
/strain="CWL029"  
/db\_xref="taxon:115713"  
99..626  
/gene="Cpn0439"  
99..626  
/gene="Cpn0439"  
/codon\_start=-1  
/transl\_table=-1  
/product="hypothetical protein"  
/protein\_id="AADI8583.1"  
/db\_xref="GI:4376722"  
/translation="MSSTLNGVPSSLPSEESADLTFTNKEIVLAGKGNVELTHSIPM  
HIAATILVVALAGIAIICLCYOSIILAVIGVLTITLLCALVAFIFIRQL  
PAAHTTVQFIREKIRPESLSIQTVAQRKTQDTLKVEELCDLSQEKRLSTLYQL  
KRELSHNEKNTQN"  
651..1289  
/gene="Cpn0440"  
651..1289  
/gene="Cpn0440"  
/codon\_start=-1  
/transl\_table=-1  
/product="hypothetical protein"  
/protein\_id="AADI8584.1"  
/db\_xref="GI:4376723"  
/translation="MATSVAPSPVPSSPLSHATEVNLFPNAVITQPHPIPAAPWET  
RSKLSKHTLCEFALTLLTLLGTISAGYAGTGMNIIICGIGLIYTLTLLALAIIP  
LKKQISTGTLIDEISODISISGFGVORGLMSTIKSVLPELTIONOKETLIMEI  
EAKESIONELKITECQNKLNQKPKRSSQSPKRSIKHLSKNPIYLDC"  
1437..2387  
/gene="Cpn0441"  
1437..2387  
/gene="Cpn0441"  
/codon\_start=-1  
/transl\_table=-1  
/product="CT007 hypothetical protein"  
/protein\_id="AADI8585.1"  
/db\_xref="GI:4376724"  
/translation="MFKLEHIAAFAGHVLSTPIFIYODAGIDEACKNPPRPESA  
OIOYLKVNDAEKKLPHTQIGYROYDGTCTCPIREHSGILSTGCIYDRIOMKST  
PISEPTPNCGAATFODTSPYNYVLSICATYTSIAKMWSTIISGIVDKNIMKSG  
LHGVLGSKQATEKLSALTEGVINETGLHDEKAMPLEVSYKATDOLLICITPVNS  
IDRSTVCNGLIAYRLTRFKRLNHLNLISSRGIEYOGREIENVKILPWPWSFIK  
GFGWSIGNDISLADHNNNKTSHTEKTSAFEGSAVMNP"  
2559..3077  
/gene="Cpn0442"  
2559..3077  
/gene="Cpn0442"  
/codon\_start=-1  
/transl\_table=-1  
/product="CT006 hypothetical protein"  
/protein\_id="AADI8586.1"  
/db\_xref="GI:4376725"  
/translation="MGEKNICKOGSOLYINGIPERLIARKLNCKAKSYPRALITIEY  
LVSSVGLAKVILPCASTYALTLPLRALFNKIKRSCHLSYAAWMLHLITLTAIV  
LILVSLVFIIPPVYFISLGLMSVTTSTVLQVHKNLPPPEPPSRPHTPPPFAD  
EYVPLISESYFD"  
3275..4528  
gene

CDS  
/gene="Cpn0443"  
3275..4528  
/gene="Cpn0443"  
/codon\_start=-1  
/transl\_table=-1  
/product="CT005 hypothetical protein"  
/protein\_id="AADI8587.1"  
/db\_xref="GI:4376726"  
/translation="MSOPPIINPLGQPVYPAASPSQSPYVRKLNKTSNGLFRFTI  
LDSPKAKKIEAITAIVPLIIVLAIAGLAFVASSGTMLVAFMFWGLITLISDI  
PSPKRYVDTGITLAIALILSLISLISGNSLMITLAPALACALGATLILSDI  
LPMSEKIDLTIDYFRREHLMEKKTQETAPRIIVTSADAKKIAVEKDKLSASAM  
EHEASORODARHRRIGREAGSGSEFYSSRRPESRSTKPKSDAASRPASI  
SPPEKDPQPHYHFKDLRSSRSGSAAFPIMPASSRSRPSSTGYLHPEYVPGG  
KEPSTIRVSSSSRRSPEDRODKOQOONODEOKQOSKKSGKSNOSLKTIPPDGKST  
ANLSPSPESDIDDERKKRHRKNK"  
4833..9056  
/gene="pmp\_6"  
/note="Cpn0444"  
4833..9056  
/gene="pmp\_6"  
/codon\_start=-1  
/transl\_table=-1  
/product="Polymorphic Outer Membrane Protein G/I Family"  
/protein\_id="AADI8588.1"  
/db\_xref="GI:4376727"  
/translation="MKSLEPLWLTSSALVPSLAPMAANTDSSDNYENGSGSNAF  
TAKETSDASGTYTTLTSDVITVNSATTPADKSCFTNTGALSFGADSHVLQITLAL  
THGGAIAINNTALSPSGFSLIDSAPATGSGKGAICVTNTEGTAFTPNASVY  
LOKNTSEKIDAAVASIDIAKTTTALDONTSTKNGALCSTANTVOGNSGVYF  
ESMTATDGGGIIYSKEDSTLDANTGVYFESMTATKNGAGMSDDNLTALNGVLEO  
ENKTGSAAGNPEGCGGAIICCYIATNTQTLAISONEMFSTMTTANGAIIYA  
TKCTIDGNTTLTFDONTATAGCGGAIITTEDSLNGSTVYFTPTATGALYSK  
GNSLTGNNTLIFSGNKATGPNSSANQDECGGAIILAFIDGSVSDTGLSIANQD  
SLTSNAAYVSGAIIYAKTCYLTGNSLTFEPGANGVSGAIIYTEDPFTLGTSGVY  
FSTNTAKTGALYSKNNLSGNTNLIFSGNKATGPNSSANQDECGGAIILFLES  
VSTRKGMIDENENSLGNTATVSGAIIYATCALHGNTLTFEDGATATAGAIYT  
ETEDPFTLGTSGTFTSTNAKAGALHTGNSFTKRNKLAYSNGATATATTQD  
EGCGGAIILCINISSDIATKSLITLSTNSLFTINNTAKRSGGIIYAPRCYSGSEST  
DGNATFSGGAIITSKNLITANCPVSTNNSGKGGAIYADSGELSLIEDIDITNS  
GLNAIBETSTPNISIHLAGAKITKLAAPHTTYFDPIEMEPASGCTIEEYINQV  
VKAIVPPQPKNGPIASVPVPAAPAPNPTIYFSSGKLSQDASIPATWTYILNQ  
INLAGGVNLTKEATLOVYSETOQDSTVEMDAGTLETTNTNDSIDILKNLSVL  
DALDGRMTLITAVNSTSGKISGDLKEHNESFYDNDELKANLNPFLIDLSSTGT  
VINDPPIPSMAAPDYGTQGSWTLPKYGAGGKTYLVAEQALGTPPELBAITY  
PNSLMAVYVNIHSIQEITATMSDABSPGIMWIGIGNAHODOKENACFRLISRY  
IVGSMTPQETTFEVAFAFSQIFGKSKDYPVSDIKSOYALCAOSSYVPLHSLSR  
HVLKYLPELGEETPLVHGQVSYGRNHHMTTRKLANTGKSDMDSHFAVEYGSLL  
PVDLNTYRLYSYSPYKLOYVSYNKGFOEVAADPRIFDASHLYVNSIIPGLTFKHS  
AKPPSALLITLGYAVDAYRDHPICLTSITNGCTSMSTPATNLSQAFAEASHKILHL  
GLDGFASGSECLNSSRSKYNANGCTRTSF"  
9318..12128  
/gene="pmp\_7"  
/note="Cpn0445"  
9318..12128  
/gene="pmp\_7"  
/codon\_start=-1  
/transl\_table=-1  
/product="Polymorphic Outer Membrane Protein G Family"  
/protein\_id="AADI8589.1"  
/db\_xref="GI:4376728"  
/translation="MKSSVWLTFPSSRISLISYAAEYTLDSNNNSYDNGSTPT  
VESTTDAAGTYSLSLSDVSFQNGALGIPLAGSGTLPEAGDITPQGNALAPFAPN  
AGSSAGTAVASTADKNLIFNDSRLSITSCPSILSPQCALKSGNLSLGNSTOI  
IFTONFSSDNGGVINTKNFLSGTSQFASRNOAFQGGVYAVAGTITINSPDI  
VSFSQNLAKSGGALYSTDNCSITDNQVLFVDSNMEAQAOGAGACCTTTPKTYVL  
TGKNNLSFTNNTALTYGALISGKLSISAGPILFOSNIGSSAGOGCGGAIINASG  
ELIASATSGDITNNQVNTGSTRAINIIDTAKYTSIRAAFGOSIYYDPTNG  
TAASDTLINLADANSEIYGAIVYSGKLSPTERAIANTYSTIRQAVYARGDL  
VLADGVYVTRQLQSPGRITMDGGTTLKACRANSLNGLAVNLSDGTNRKALKT  
EADKNLSLGTALIDTEGSFYENHLKASSTYPLLELTAAAGTITLGAISLTL  
QEPETHYGOGNMOLSWANATSSKIGSINTTRGCIYSPSRKSNLPLNSLMGFIDIR  
SINQILTKSSGPEFERELWLSGIANFYDMSPTRHGRHISGVALGITATTPAD



[illegible][illegible]

gene  
CDS  
/gene="CPJ0271"  
/codon\_start=1  
/transl\_table=11  
/product="lysophospholipase esterase"  
/protein\_id="BA98481.1"  
/db\_xref="GI:8978645"  
/transl\_table="MTDYSFFERRKIGNIEAIECPONDPPIIILCHGYSIADNLPFFEE  
PSIGSEFSKRLPWPWIPNGILPLENDRSRACPPLNVLLOESILVANGVGNLOEYK  
DLFPVDLETPKEALEELILNINRPYNTIIGCSOGIILATHVITSONGYACALIRY  
AGARLFCNOMEGSLKQCAOYPLQSHGIEDELILPHLGAHLNDLLITRLNQFVSHFG  
GHEIISVYFQKQIVTFPMIDPARG  
complement(4359) .5255)  
/gene="dnax\_2"  
complement(4359) .5255)  
/gene="dnax\_2"  
/codon\_start=1  
/transl\_table=11  
/product="DNA polymerase III gamma and tau"  
/protein\_id="BA98482.1"  
/db\_xref="GI:8978646"  
/transl\_table="MHLEENQGEALRLKRYHOEVPAILHGHFTLPVLQDAEQOLA  
YIILHADDRTTLTAISAFILKVEPEPPKHAIVILITAKORLITISLSITSEIRGKK  
TLCSEFYSYLFERYACQELPTEVSOIIRKESSEFDKQYLNRKQVRPHEVLEILRDXY  
TILSLCKASALYLPHEVKEITLQPLPLIDKVLILVESACRSLNSSASVLEWVAIQ  
LVSIYQKKEKELVSVSGQDLSN"  
complement(5246) .5866)  
/gene="tdk"  
complement(5246) .5866)  
/gene="tdk"  
complement(5246) .5866)  
/codon\_start=1  
/transl\_table=11  
/product="thymidylate kinase"  
/protein\_id="BA98483.1"  
/db\_xref="GI:8978647"  
/transl\_table="MEVIEGEGSGSKSLAKGDLQVAODRKVLLTRBPGGLIIE  
RLRIILBPEHLEISRCCELPLFGSRGSLIDQVITLPGQYIVICERFHDSTVIOE  
GIAELGLADPVAIDCKRYGCPPLPNVLLDIPADIGLQRRKQRYFDKEFKPLIS  
YHNRRREGSLASADPSRYLVLDARBSLASLIDKVMILHTGLCT"  
complement(5871) .8375)  
/gene="gyra\_1"  
complement(5871) .8375)  
/gene="gyra\_1"  
complement(5871) .8375)  
/gene="gyra\_1"  
complement(5871) .8375)  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase subunit A"  
/protein\_id="BA98484.1"  
/db\_xref="GI:8978648"  
/transl\_table="MFNKDELIVPRNLEENKESYLRYSMSVILSRALPDIRDGLKPS  
ORVLYLPTISQILSPGAKHKRCAKICGQTSQDYPHGSGSYVPLVMAQKMAWRYPL  
VYDQGNFNSIDDDPPAARMYTEARLTSTAMLMIDLDQTDVDPYANDEQMAWRYPL  
SKFPLILNGSSGIALVGMATNIPHNIGELIEALLILLANQA SVDESLIOMRDPDP  
TGCIITCSGSEIRSTYTTTRGKIKYARARLHYVENDKDHRESITITEMPYNNKSLIIO  
IANLVNENTVLGISDVRSDSKDIRVYLETKGSESEIILNRYKFTDVQVYFGAM  
LALNDLPTPMSIHRMISAMIRHREYLRRTRELKNAETRAVNLGELYKALSCDLA  
LVKTLREGNEEHAKERIIESEGFTEPOLAILRLYLQGLAEKLTQKEYEELNKL  
IAYYQVSDGLVADDIRNELODLKHKYARRTTIEPDADRIDIEDIITNNSYIIT  
TISGDYDKRKDPVYKFORGRGHGQYGFQMKKGAGTLKAYSAFTQDYLITPNEQO  
CYWLKVMQLPGEERRAKKPIILNLEIGIRPEEELAILINKFNDAQFLFLATKRGVY  
KRVISDAENSRPKKGIIRALELDEDELLAACHIVSDEKWLFTHLGAARFPHEKRY  
PMGRFARVCRVSKLNEDKRYSOIYVENSQVILVDCDQFGKRSILVEDERETNRGV  
GVSRLILBERNGVNYGAIPYVDHDSIILMSQCAIRINMDVAVMGRSQGVRLVHL  
REGDLVLMETLSSNENDEVLVDSGSEECQPTVSLR"  
complement(8391) .10808)  
/gene="gyrB\_1"  
complement(8391) .10808)  
/gene="gyrB\_1"  
complement(8391) .10808)  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase subunit B"  
/protein\_id="BA98485.1"  
/db\_xref="GI:8978649"  
/transl\_table="MDPEKKNYDAISATVLEGLQAVREBRQMYIGDYITGLHLLHYE

	Query Match	Best Local Similarity	Score	Pred. No.	Length
	Matches 1055; Conservative 0;	Mismatches 1032; Indels 75; Gaps 10;			
gene					
CDS					
gene	complement(10812..11159)				
CDS	/gene="CPj0276"				
	complement(10812..11159)				
	/gene="CPj0276"				
	/codon_start=1				
	/transl_table=11				
	/product="CT191 hypothetical protein"				
	/protein_id="BA96486.1"				
	/db_xref="gi:8978650"				
	/translation="MFLKRRKKGSQYONKRRTASPIKHAKHYLVNLOELOKIARAARHDAIDAMNOVERDRKYKMSQAIGFRDHILVLYKNSSYLALKQTPODMLMSLYQVA SHVOIREPLFG"				
Query Match	8.5%; Score 260.8; DB 2; Length 300650;				
Best Local Similarity	48.8%; Pred. No. 2.3e-52;				
Matches 1055; Conservative 0;	Mismatches 1032; Indels 75; Gaps 10;				
Oy	657 GAATGCTTCGTGAGGGCGTCTTGACTCCGATGGTATATTGAATTATGCATAACAAATGCTTA	716			
Db	195166 GAAGAAGCTTTGGCGGTGCTGTGTACAGCAGTACAACATGTTGCTGATTAAGATAAATTTTCA	195225			
Oy	717 TGTTCATTATTCAGAAAAATFAGACATTCACATCTCATCTFATAGGTAAGGAGGCGCTGTCTG	776			
Db	195226 AGTAGATCTTTGACGCCAATAGTGCTTGGGAAGCCGCTCAAGCTCAGGGCGGCTATTITGG	195285			
Oy	777 TTGTCTCCCATCTGAGGAAGTAGTACTCCAGTTCCTATGTGACTTTCCTGCACAATPA	836			
Db	195286 TTGC-----ACTACGACAGATPAAAACAGTCACTCTCACGGAACAA	195327			
Oy	837 ACAGTTACTCTTTGAAGAAACCATTCCATATATGGTGGCGGAGGCATTATAGCTAGAA	896			
Db	195328 AAACTCTCTTTACAAATAATACAGCATFTGCATATATGGCGGCATCTCTGCAGCTCAA	195387			
Oy	897 ACTTAGCATCTCTTCAGGAGGCTCACACIAATTTATATATATATATATATGCAAAATTC	956			
Db	195388 GGTCAGTTTTTCOCGTGGAGGCTCACTCTATTTTAAAGTATATATCTCAGGAAGTAGGCC	195447			
Oy	957 ---GCAAAAATTTAGGTGAGCTATTGCCATTGATACGTGAGGAGGAGATCAGTTATTACGC	1013			
Db	195448 CGGTACGAGGAGGAGGAGGAGCATCAATATATGACATCTGCTGGGAACCTCGCTCTGTC	195507			
Oy	1014 AGAGAAAGAACATTAAT-----CCAAGGAAACCGGACGACTTACGCTTTT	1064			
Db	195508 TACTTCTGGAGATATATACCTTCATTAACACCAACGATCCACAAACGAGACACAAATGACAAG	195567			
Oy	1065 GAATGGCATCATCTTTTACAAAAATGCTAAATTCCTGAAATTTACAGGGGAGAAATGGATG	1124			
Db	195568 AAACCAATTAATATCATATTGATACCGCTTAAGTGCATCATGATCAGAGTCTTAACGGGCA	195627			
Oy	1125 CTTCTAGAAATTTTATATCTTATCTTGAAGCAGATGGGTCTTACC-----	1174			
Db	195628 ATCTATCTATTTTCTATGTATCCCATACAAATTCAGGAACCGCAGCTTTACCGACACATTT	195687			
Oy	1175 --AATGAATATCAACGAGATCTTAAAAATTAAGAGTACACAGGAGACCATCTTTTC	1232			
Db	195688 GAACCTTAACCTTAGACAGATGGGAACAGTAGAGATATGGGGGTGGCATTTCTTTTC	195747			
Oy	1233 TGGAGAAAAAGCTACGAAACGA-----TCTTAGGGATTTTAAATCTACATCCC	1283			



Db 195748 TGGAGAAAAGCTTTCCTCCCTACAGAAAAAGCAATGCGTCAACGTCACCTCTACTATCCG 195807  
 QY 1284 TCAGACGCTCAACCTGCTGCGAGATACCTTATTAAGAGGGGGCCGAAGTCACAGT 1343  
 Db 195808 ACAACCTGCACTATTAGGCGGGAGATCTTGACTCTGATGAGAGCACCGACTT 195867  
 QY 1344 TTCAAAATTCACGAGTCCGAGATGCAATTTAGTTTATATTAGAACCAATCAT 1403  
 Db 195868 CAAGATCTGCTCAAGATCCAGATCCCGATCTTAATGATGGGGGACTACACTTAG 195927  
 QY 1404 AGCTCTAAGAGACATTTGCCATCACAGGCGCTCGCATAGATATATAGACTTAAAGTTC 1463  
 Db 195928 TGCTAAGAGCAATCTTTCGCTTAATGGCTTAAATCTCTCTCTCTTATAGATGG 195987  
 QY 1464 ATCTCAACGACCTGTTATTAAGCAACCGCAATTAAGATATCCGTCAGGGA 1533  
 Db 195888 AACCAACAGGACCTTTT---AAAACAGAGCTGAGATATAAATATCACACCTATCGGG 196044  
 QY 1524 CTCATAGAACTTATCTGCGCTACTGCAATGCTATGAGATCTCAGAAATGAGAAATTC 1583  
 Db 196045 AAGCATTCGCGCTTATTAACAGGAAAGGTCATCTATGAGATATATTAATAAGTGC 196104  
 QY 1584 ACAGAGCTTCCCTCTGCT-----CTCTTAGAGCGCTGAGCCGGGGGTAGTGT 1631  
 Db 196105 TAGTACCTATCTCTCTTGAACCTTACACCGAGAGCCAAAGGAAGATTAATCTGGG 196164  
 QY 1632 GACTGTAACTGCTGAGATTTCCATCCGGTAAAGTCCCATTAATGCTTTCAAGCAATTC 1631  
 Db 196165 AGCTCTTCAACCTGACTCTTCAAGAACTGAAACCCACACGCGGTATCAAGAAACTG 196224  
 QY 1692 GAAATAGCTTGACACA---GGAAGTGAACAAAGTTGAGAAATTCCTGCGATTAAT 1748  
 Db 196225 GCACTTCTCTGGGCAATGCAACATCTCAAAATATAGAGCATCAACTGACACCGCTAC 196284  
 QY 1749 AAATTAAGCTTACCTGAGAAAGAAAGAAATTAATTAATTAATTAATTAATTAATTAAT 1808  
 Db 196285 AGATTAATCTCTAGCTGAGAGAAAGAAATTAATTAATTAATTAATTAATTAATTAAT 196344  
 QY 1809 TGCTGTAAAGTGCAGATCTTAATGAGAGTTCAGAGAACCCATGATGAGCTTACAGAC 1868  
 Db 196345 CTTTATGATATACGCTGATCAATGAGCTTATAGAAACCAAGTCCAGGGGAGCTTT 196404  
 QY 1869 AGATGAGGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1928  
 Db 196405 TGAGCGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 196464  
 QY 1929 CAATTAAGTACCGCTATTAACAGCGGTGATGATGATGATGATGATGATGATGATGATGAT 1968  
 Db 196465 CCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 196524  
 QY 1989 ACCTAAGCACTATACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2048  
 Db 196525 TGCGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 196584  
 QY 2049 GGTTCACAAACGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2108  
 Db 196585 TACAGGTAACACGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 196644  
 QY 2109 CTAAGGAAATATTTTCGCTTATGCTGCTTAACCTTAATGATTAACCTGCGGATTCCTC 2168  
 Db 196645 GCTCTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 196704  
 QY 2169 AAGAGGTTCTTCAAAATCTCTTATGATTTTCTATTTTGTGCTTATGATGATGATGATGATGAT 2228  
 Db 196705 TGAATCTCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 196761  
 QY 2229 CACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2288  
 Db 196762 AGACAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 196821  
 QY 2289 CAATGTTGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2348  
 Db 196822 CGATGCTTCTGTCAGATCTTGGAGCTACGCTGCGT---TTGTATATTTCCGTTCCGTA 196878

QY 2349 ACTTTCAGAGTGCATCCATTTATGAACCTACATTAATTAATGATGATGATGATGATGATGATGAT 2408  
 Db 196879 TCTTCTGAAGAAAGTCAAGCTTTTGTCAAGTACAGTATATATATATATATATATATATATATATAT 196938  
 QY 2409 TTTCAGAGACAGTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2468  
 Db 196939 CTTTACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 196998  
 QY 2469 TGATCTTGAAGTACGCTTTGAGAAAGCTGACCTTCTCAGATGATGATGATGATGATGATGATGAT 2528  
 Db 196999 GATTCCTATAGGCTGACCTTCAAGAGAGCTCAAAAATCAAAAAAGGAACTTACAGATCT 197058  
 QY 2529 TAGTCTCTTAT 2588  
 Db 197059 TACTTAT 197118  
 QY 2589 GATTAAGGAGACCTGCTGCTTGTCCGGCAGCAGACAGTATCAAGACATGCTTTTGTAGG 2648  
 Db 197119 AGCTAGGATCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 197178  
 QY 2649 GAGTGAACGGGTGATACCTTAACGATATACGATGATGATGATGATGATGATGATGATGATGAT 2708  
 Db 197179 TCGTCTGCGAAGCATTTCCAAAGTAAACCCACATGGAATCTTGGGTCAATTCGCTT 197238  
 QY 2709 AGAATGCGCGCCCATGCTAGGAATTAATATATATATATATATATATATATATATATATATAT 2768  
 Db 197239 TGAAGTACGAATTTCTTACGAAATTAATATATATATATATATATATATATATATATATAT 197298  
 QY 2769 GA 2770  
 Db 197299 GA 197300

RESULT 14  
 LOCUS AB1827 3200 bp DNA PAT 21-JAN-2000  
 DEFINITION Sequence 1 from Patent WO9858953.  
 ACCESSION AB1827 GI:6731863  
 VERSION AB1827.1 GI:6731863  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 3200)  
 AUTHORS Madsen, A. and Birkelund, S.  
 TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHILAMYDIA PNEUMONIAE  
 JOURNAL Patent: WO 9858953-A 1 30-DEC-1998;  
 MADSEN ANNA SOPHE (DK); BIRKELUND SVEND (DK)  
 FEATURES  
 source location/Qualifiers  
 1..3200  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 205..>2987  
 /note="unmanned protein product"  
 /codon\_start=1  
 /protein\_id="CAB69199.1"  
 /db\_xref="GI:6731864"  
 /translation="MRTSLIPWVYVSVLAFSCHLQSLANELLSPDSEFNGNIDSGTF  
 TPTSATVYSLNGDVEFEYERKGTPLSDSCFRTDNLPLFGLNGHSLTFEGIDAGNHA  
 GAASTYKANKNLTPSGFSLSPDSSTVYTNQGLTSSAGVNLNIRKLYVAGNFS  
 TADGAIKASFLITGSGDALFSNNSSTKGGAIATTKAGRIANNIGVYRFLSNLAS  
 TSGCAIDDEGSLTSLNNKPLFEFGNAKATGGAIICWTKASGPELITSNKTLIFASN  
 VAEISGGAIAHAKKLALSOGTEFLERNNVSSAIPKGAISIDASBELISAETGNITF  
 VRNTLLTTSSTDPKRNAINIGNGFTEFLRAKNHTIFEDYDITSEGTSSDVLKIN  
 GSAGALNPYOGTILFSGETLADBLVADNLKSFOPVPSLGGKLLQKGVLEST  
 FSGEAGSLIGMOSGTLTSTAGITITNGIIVDSIGLQOPVSLTAKGASNKYIVSGK  
 LNIIDIEGNIVESHMSHDLQSLIKITVDADVDVWDVSSILIPVADPNSEYFGOC  
 QNRVNTTPTATATKAT  
 ATGMEHKGQFVWSSMTNPLAKTGDENRKGFRHTSGGVYIGSAHTPRKDLFTFAFCHL  
 FARDKDCFIAHNNSRTYGGTLPFKHSHTLQOPVYLRGAKFSESAIERFPRIPAL  
 DVQVSRSHDNRMEHTYSLPESEGSWSNECIAGIGLIDPLVLSNHPLEKTFIDOM  
 KYEMVYVQNSPFESSDDKGFSLIGRLNLISIVGAKFVGQDIDGSDYTDLSGFVSD

## COS





OY	2070	AATGATATTAGATCGATATCTATCAATATACAAACCCCTAGGAAATATTTCCGTTA	2129	TITLE	Identification of two novel genes encoding 97- to 99-kilodalton
Db	2295	AACCTAGGAGGAGCACTTATTTCTCAAGCACTCTCATACCTACACCCCAAACTTTT	2354	JOURNAL	outer membrane proteins of Chlamydia pneumoniae
OY	2130	TGCTTCGCGTAACCCATATGTAACGTCGGGATTCTCTCAAGAGTTTCTTCAAAATCC	2189	MEDLINE	Infect. Immun. 67 (1), 375-383 (1999)
Db	2355	GAGATATGGAAGACAAAGTTTCTGATATCAGTATGAAAAATTTCCCTAGGAAATTC	2414	FEATURES	
OY	2190	TCTATATATTTTTCATTTTGTGTGCTATGTCATGCCAATATATGAAAAACA	2249	source	location/Qualifiers
Db	2415	CCGA--GCCTTGATGTCCAAAGTTTGTTCAGCATTCAGACACCGTATGGAACGCA	2471		1..6030
OY	2250	CTACGCAATTTTCCATTTGGTGAAGAAACAGCTGGAACAAATTTGGCTATAGAGTG	2309		/organism="Chlamydia pneumoniae"
Db	2472	CTATACCTCATTTGCGCAATCCGAGGTTCTTGAGCAACAGATGTAATAGTGTGAT	2531		/isolate="CDC/CWL-029/VR-1310"
OY	2310	CGAGGAGGAGCATGCTCTATTTGATTTGAGAACGGAAGACTTTTCCAGGTGCCATCC	2369	gene	/db_xref="taxon:83558"
Db	2532	CGGCTAGACCTTCTTTGTTTCCAAACCCATCCTTTTCAAGACCTTCATTC	2591	CDS	/clone="pEX2-18"
OY	2370	ATTATATCAACATCAATTAATGCTTATCGAGGAGATTCAAGAGACGACTGCAGA	2429		/clone="pB8F3"
Db	2592	ACAGATGAAAGTGAATGTTATGATACAAATAGCTTTCGAAAGCTCTAGTGA	2651		/clone_1lb="pbluescript"
OY	2430	TGGCCGTAGATTTAGTATAGGAGTTTACATGATGATTTCTGACCTAGCATACGCTT	2489		/clone_1lb="pEX2"
Db	2652	TGGCCGTAGATTTAGTATAGGAGTTTACATGATGATTTCTGACCTAGCATACGCTT	2711		/dev_stage="elementary body"
OY	2490	TGAGAACGTGCACTTCTCAGATGATCTATGACTTATGATTTCTATATATCTCTGA	2549		complement(1..2804)
Db	2712	CGTACAGGGGATATCGGAATTCCTACACCTATGATCTCTCAGAGATTTCTTGTTCGA	2771		/gene="omp5"
OY	2550	TATTTTCCGTAGATGATCCCTCATGATGAGTGCCTCTGATTTAGCGGAGACTCTGCT	2699		/note="putative lipoprotein"
Db	2772	TGCTATCGTAACATCCCAATCTACAGGACTCTTATGAGCCCAACTCTTGTGA	2831		/codon_start=1
OY	2610	TGTTCCGCGACACACATCATGATGATGAGGAGTGAACGGGTGCGTATCA	2669		/transl_table=11
Db	2832	AATTCGGGGGCAATTTCTCAAGACGATTTTACTGAGGGGTGACAAACACTACGT	2891		/product="outer membrane protein 5"
OY	2670	CTTTAGGACTATAGTACTGATTTATGTCGAGAAATGATGAGCCGCCCATGCTAG	2729		/protein_id="CAA04671.1"
Db	2892	CTACACTCAATGTGAGCTCTCGGACATTTACGATGGAACCTCGGTGATCTCAAG	2951	mat_peptide	/db_xref="SPRMBL:086163"
OY	2730	GAATTAATAATAACGTGGAAGCAATTTCTTTAGAGTTTCAATGCTGTGT	2789	sig_peptide	/translation="MKSQPSMLVLSSTLACFTGCTVFAATATENIGPSDFDSSTWNC
Db	2952	GAATTAATAATAACGTGGAAGCAATTTCTTTAGAGTTTCAATGCTGTGT	3011	RBS	TYPTKNTTGTIDTTLTGDTITLQMLGDSALATKCFSDTTESSLFAGKGLSLSLNLIK
OY	2790	GGTCCGATCTTACT	2806		SABGAALSVTTDKNLSTGESSLTFLEAPSVTTTPSGKAVCGGDLFFDNNGTILF
Db	3012	CTTCTAGGAGATTTC	3028		KODYCEENGALSTKMLSLKNSYSSIFEBNKSASAKGAGCAICAGTVDITNNAPT
RESULT 15					LFENNTAEAGGAINSTGCTTGATGATGATGATGATGATGATGATGATGATGATGAT
LOCUS	CPOMP54	6030 bp	DNA		VFPSGQAVANGAATYAKKLLASGGGGISFENNIVOGTTAGNGAASILAAGECSL
DEFINITION	Chlamydia pneumoniae omp5 and omp4 genes.				SARAGDITPENGNAIVATTPQTTKRNISIDIGSTAKITINLARSISGHISFEFDITANRL
ACCESSION	AF001311.1	GI:3255934			DSDITLANKADGNSDTSISIVFSEGEKISEDEAAVADNLSTLQPTLYTLRGNLYL
VERSION	AJ001311.1	GI:3255934			KRGVTLDTGFTGTAGSVIMDAGTLKASTETVTLGSLIPVDSLEBGRKVIYASPA
KEYWORDS	omp4 gene; omp5 gene; outer membrane protein 4; outer membrane protein 5.				ASKNVALSGPILLDLNOGNAAYENHDLGKTQDFSEVOL.SALGTATTTDVPAPVAPPT
SOURCE	Chlamydia pneumoniae.				HYGYOCTGWTWVDPTASTPKTATLATLNTGYLNPBEROGPLVPSNLGFSFSDIA
ORGANISM	Bacteria; Chlamydiales; Chlamydiaeae; Chlamydia.				IOGVIRSATLTCSDRGFMAAGVAFNLADKDKGKRRYRHKSGGYAIGGAAGTSCSEU
REFERENCE	1 (bases 1 to 6030)				ISFAFOLGSDGDFLVAKNHPDTYGAFTYOHITGCSGFIGLIDKLPESWGHKPLV
TITLE	Direct Submission				LEQQLAYSHVSNLKTRTYRYPVYKSGSNNAANMLGASHSHPYELHREFDYAPYI
JOURNAL	Submitted (29-AUG-1997) Knudsen K., Department of Medical				KLVLTIRDSFSEKTEGKSFDSNLFNLSLDTGVKFEKFSQNDSDTLTSLYPD
AUTHORS	Microbiology and Immunology, University of Aarhus, The Bartholin				LIRNDRCTTALVIGASWETVANNLRALQVRAQSHYAFSPMEVLQGFVEVGS
REFERENCE	2 (bases 1 to 6030)				
AUTHORS	Knudsen, K., Madsen, A.S., Mygind, P., Christiansen, G. and				
	Birkelund, S.				



Db 4635 TACAATATACAAAGAGGCCAGGCACTTGGACCAAAAACAGATTGTTCACAGCCCGA 4694  
QY 1770 AAAAGAGGAATTTAGTCTCTAATATCTTGTGGGGAAAGCTGTAAATGTCAGATCCTT 1829  
Db 4695 AAGAAATCTCGTTAGTATGCAATACCTTATGGGAGTCTTTTACTGACATTCCTCTCT 4754  
QY 1830 AATGAGGTTCAAGAGACCCATGATGAGCTTACAGACAGATCGAGGGCTGTGATCGA 1889  
Db 4755 GCAACACCTTGTAGAGATCGGCACTGGATGTGAAACACAAACAGGTTTCTGGGTTTC 4814  
QY 1890 TGAATTTGGGAATTTCTTCATGTATCTGCCCTCCAGACAAATATTAAGTACCTCATTA 1949  
Db 4815 CTCATGACGACACTTCCCTGATAGACTGGAGATGAAATGCAAGGCTTCCGTCATAC 4874  
QY 1950 CAGCGGTGATATGTCTATCTGTAATATAGATACACACTAAGCATATACCTTGAT 2009  
Db 4875 CTCTGAGAGGCTACATCGGTGAGAGTGCACACCTCTAAGACGACCTATTACCTT 4934  
QY 2010 GGCATTTTCCCACTCTTATAGACAAAGCACTATGCGGTTTCCAAACAGAAATACAG 2069  
Db 4935 TGGCTTCTGCTCTCTTGTCTAGACAAAGATGTTTATCCCTCACAACTCTAG 4994  
QY 2070 AATGTATTTAGATCTATCTCTATCAATATACAACTCCCTAGGAAATATTTCCGTTA 2129  
Db 4995 AACCTAGGATGAGACTTATTTCTCAAGCACTCATACCTTACACCCCAAACTATTT 5054  
QY 2130 TGTTCGCGTACCCCTAATGTAAGTCGGGATTCCTCAAGAAAGTTTCTTCAAAATCC 2189  
Db 5055 GAGATTTAGAAAGCAAGATTTCTGAATCAGTATGAAAAATTCCTAGGAAATTC 5114  
QY 2190 TCTATATATTTTATTTTGTGTGCTATGCTATGCCCAATGATATGAACAGAGA 2249  
Db 5115 CCTA--GCCTTGATGTCCAAAGTTTGTTCAGCCATTCAGACACCGTATGGAACGCA 5171  
QY 2250 CTACGCAATTTCCCTATGTTGTAAGAAACAGCTGAGAAACAAATTTGGGCTATAGAGTG 2309  
Db 5172 CTATACCTCATGTCGAATCCGAAGTCTTGTGAGACAGAGTGTATAGCTGGTGTAT 5231  
QY 2310 CGAGAGGAGCATGCTCTATTTGTAATTTGAGAACGAAAGACTTTTCCAAAGTGCATCCC 2369  
Db 5232 CGGCTAGACCTCTTGTGTTGTTCCAAACCCACATCTCTTTTCAAGACCTTCATTC 5291  
QY 2370 ATTATGAACTACATATGATTTATGCTATACAGGAGATTTCAAGAGACGACTGAGA 2429  
Db 5292 ACAGATGAAAGTCGAAATGTTATGTATCACAAATATAGCTTTCGAAAGCTCTAGTGA 5351  
QY 2430 TGGCGTAGATTTAGTATGAGAGTTTAACATCGATTTCTGTACTGTAGCATACGCTT 2489  
Db 5352 TGGCGTGTGTTTGTATGTGAAGGCTGTAACTCTCGATTCGTGTGGGTGCGAAATT 5411  
QY 2490 TGAGAGCTGGCATTCTCTAGATGATCTATGACTTATAGTTTCTCTATATTCCTGA 2549  
Db 5412 CGTGACAGGGGATTTGCGAGATTCCTACACCTATGATCTCAGAGATTTCTGTTCCGA 5471  
QY 2550 TATTTTCGTAAGGATCCCTCATGATGAAAGCTGTGATGATGAGCGAGACCTCGCT 2609  
Db 5472 TGTCTATGTACATCCCATCTACAGCAGCTTGTGTAGAGCCCGACCTCTTGGA 5531  
QY 2610 TGTTCGCGAGCACATCATACAGACATGCTTTGTAGGAGTGAACGGGTGCGTATCA 2669  
Db 5532 AATTTCGGGATGCGCAATCTTCAAGAGGCAATTTTACTGAGGGTGAACAACTACGCT 5591  
QY 2670 CTTTACGACTATAGCTCTTATGTGAGAGATATAGATGCGCCCATGCTAG 2729  
Db 5592 CTACAACTCCAAATGTAGCTCTCGGACATTTACGATATGGAACCTCGTGTCTCAAG 5651  
QY 2730 GAATTAATTAATTAACGTGGAACAAATTCGTTTAAAGSTTTCATTCGCTGTGT 2789  
Db 5652 GAACCTACATGTATGATGTTGTACCAAACTCCGATCTAGATTGCTAAAACTCCCTAGTT 5711  
QY 2790 GGTTCGGATCTTAAT 2806  
Db 5712 CTCTAGGAGGTTTCT 5728

Search completed: October 2, 2001, 00:58:17  
Job time: 44170 sec

**THIS PAGE BLANK (USPTO)**

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 04:35:26 ; Search time 800.41 Seconds  
(without alignments)  
2394.218 Million cell updates/sec

Title: US-09-446-677B-5

Perfect score: 3052  
Sequence: 1 ATGGGATTTCGCTCTGCGG.....GTTCCTTGAGTCCTTGCA 3052

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_0601:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3052	100.0	3052	20	AA06818 Chlamydia pneumoni
2	3037.8	99.5	1230025	20	AA06818 Nucleotide sequenc
3	2843.6	93.2	2950	21	AA06818 DNA encoding Chlam
4	1192	39.1	1200	20	AA06818 Chlamydia pneumoni
5	273.2	9.0	3000	21	AA06818 Chlamydia pneumoni
6	270	8.8	2787	20	AA06818 Chlamydia pneumoni
7	260.8	8.5	3050	21	AA06818 Chlamydia pneumoni
8	259.2	8.5	273254	21	AA06818 Chlamydia pneumoni
9	252.6	8.3	2950	21	AA06818 Chlamydia pneumoni
10	252.6	8.3	3200	20	AA06818 Chlamydia pneumoni
11	242.2	7.9	2808	21	AA06818 Chlamydia pneumoni

12	235.4	7.7	2784	21	AAA30848 Chlamydia antigen
13	232.6	7.6	273254	21	AA06818 Chlamydia pneumoni
14	231.6	7.6	2815	20	AA06818 Chlamydia pneumoni
15	228.2	7.5	3000	21	AAA30853 Chlamydia antigen
16	226.6	7.4	2793	20	AA06818 Chlamydia pneumoni
17	223.2	7.3	2950	21	AAA30849 Chlamydia antigen
18	222.8	7.3	3000	20	AA06818 Chlamydia pneumoni
19	220.6	7.2	2957	21	AA06818 Chlamydia antigen
20	217.4	7.1	2757	20	AA06818 Chlamydia pneumoni
21	208.4	6.8	2790	21	AAA30854 Chlamydia antigen
22	198.4	6.4	2784	21	AAA30850 Chlamydia antigen
23	189.8	6.2	2950	21	AAA30851 Chlamydia antigen
24	186.2	6.1	3050	21	AA06818 Chlamydia antigen
25	184.6	6.0	2787	20	AA06818 Chlamydia antigen
26	183.2	6.0	2784	21	AAA30852 Chlamydia antigen
27	178	5.8	2838	20	AA06818 Chlamydia pneumoni
28	159.4	5.2	2750	21	AA06818 Chlamydia antigen
29	156.8	5.1	2526	21	AA06818 Chlamydia antigen
30	150.4	4.9	2526	20	AA06818 Chlamydia antigen
31	150	4.9	840	20	AA06818 Chlamydia antigen
32	107.6	3.5	3150	21	AA06818 Chlamydia antigen
33	105.2	3.4	936	22	AA06818 Chlamydia antigen
34	105.2	3.4	936	22	AA06818 Chlamydia antigen
35	105.2	3.4	936	22	AA06818 Chlamydia antigen
36	105.2	3.4	936	22	AA06818 Chlamydia antigen
37	105.2	3.4	936	22	AA06818 Chlamydia antigen
38	105.2	3.4	936	22	AA06818 Chlamydia antigen
39	105	3.4	1830	20	AA06818 Chlamydia antigen
40	103.2	3.4	936	22	AA06818 Chlamydia antigen
41	103.2	3.4	936	22	AA06818 Chlamydia antigen
42	103.2	3.4	936	22	AA06818 Chlamydia antigen
43	103.2	3.4	936	22	AA06818 Chlamydia antigen
44	103.2	3.4	936	22	AA06818 Chlamydia antigen
45	103.2	3.4	936	22	AA06818 Chlamydia antigen

#### ALIGNMENTS

RESULT 1	
ID	AA06818 standard; DNA: 3052 BP.
AC	AA06818:
XX	
XX	26-APR-1999 (first entry)
XX	Chlamydia pneumoniae surface exposed protein Omp6 DNA.
XX	Omp6: outer membrane protein 6; surface exposed protein; antigen;
KW	Infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX	
OS	Chlamydia pneumoniae.
XX	
XX	Key
FT	CDS
FT	Location/Qualifiers
XX	1..2764
XX	/tag a
XX	
XX	W09858953-A2.
XX	
XX	30-DEC-1998.
XX	
XX	19-JUN-1998: 98MO-DK00266.
XX	
XX	23-JUN-1997: 97DK-0000744.
XX	
XX	(BIRK/) BIRKELUND S.
XX	(CHR/) CHRISTIANSEN G.
XX	
XX	Birkelund S, Christiansen G, Knudsen K, Madsen A.
XX	Mygind P.
XX	
XX	WPI: 1999-105610/09.

DR P-PSDB; AAW88419.

XX Species-specific test for identifying mammals infected with  
PT Chlamydia pneumoniae - comprises detecting antibodies specific for  
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding  
PS these proteins  
XX

PS Claim 6; Page 45-46; 115pp; English.

XX This DNA sequence codes for the novel 100.3 kDa surface exposed  
CC protein Omp6 (see AAW88419) of the human respiratory pathogen  
CC Chlamydia pneumoniae. By generating antibodies against C.  
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)  
CC was obtained which reacted with outer membrane proteins. The  
CC antibody was used to identify the genes (see AAX06816-27) encoding  
CC Omp4-Omp5 proteins (see AAW88417-28) in an expression library of  
CC pneumoniae DNA. The genes are situated in 2 gene clusters:  
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in  
CC the other, and encode polypeptides of about 89.6-100.3 kDa and  
CC about 56.1 kDa. The invention provides a new species specific test  
CC for identifying mammals (including humans) infected with Chlamydia  
CC pneumoniae. The test comprises detecting antibodies specific for  
CC Omp4-Omp5 or detecting nucleic acid fragments encoding these outer  
CC membrane proteins, especially by PCR. The proteins are also used  
CC in the diagnosis of C. pneumoniae infection in mammals. The  
CC nucleic acids and proteins can also be used in the immunization of  
CC mammals, the nucleic acids being particularly useful as DNA  
CC vaccines for effecting in vivo expression of antigens. The  
CC vaccines may also prevent atherosclerosis and bronchial asthma,  
CC which are possibly associated with C. pneumoniae.

XX Sequence 3052 BP; 875 A; 603 C; 653 G; 921 T; 0 other;

Query Match 100.0%; Score 3052; DB 20; Length 3052;  
Best local similarity 100.0%; Pred. No. 0;

Matches 3052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCCGATTTTCGCTGCGGATTTCCCTAGTCTTCTTCTTAAACATGCTCAGTCTTC 60  
DB 1 atcgatcttcgcctcgcgagatttcctcctagtttcttctaaccattgcctcaagcttc 60  
QY 61 GACATCTTCTTGAAGTGTCTACGATTTCTTAAACCCAGAGATGTTTCATGAGAT 120  
DB 61 gacattctcttggtgctactacgattctttaaaccagaagatagtttcatgagat 120  
QY 121 AGTCAGATTCGACAGCTGCTTATTAATGTCACGCTGGGATGCTATACCTTACGCT 180  
DB 121 agtcagaatgcagaaagcttcttaaatgctcaagctggagatgctactagctggt 180  
QY 181 GATGTCATATCTAATCAGTCGATTAATCTGATTAATTAAGCTGCTCAATGTGAC 240  
DB 181 gatgtctcaatatctaaatcagtcgataactcgtccttaataaaccgcttcaatgtagc 240  
QY 241 TCAGGAAGTGTGAGTTCGACGAAATCATCAGGTTATATTTAATAATATTTCTCA 300  
DB 241 tcaggaagtgtagcgttcgagaatacatcagtggttataatttaataatattctca 300  
QY 301 GGAAGCTACAAAGGAGGGGTGACTTTGTTGCCAAGATCCTCAAGAACGAGCGTTT 360  
DB 301 ggaactacaaaggaagggtgactttgttgccaagatcccaagaagcagcagctt 360  
QY 361 TCTGGGTTCTCCACGCTCTCTTATTTACAGAGCCCGAGATATTAAAGAGGAGTGT 420  
DB 361 tctgggttctccacgctctcttatttcagagcccgagatattaaagcagggatg 420  
QY 421 CTCTATTTCAAAAAATGCACTTATGCTTTAAACAATTATGATGCGTTTGAACAAAC 480  
DB 421 ctctatctcaaaaaatgcaatgcttcttaacaattatgtagtgcgttttgacaaac 480  
QY 481 CAAGTAAGACTAAAGCGAGCTATTAGTGGGCGCAATTTACTATAGAGCACTAC 540  
DB 481 caagtaagactaaagcgagctattagtgggcgaaatgtagtactatagtagcaactac 540

QY 541 GATTCGCTCTTTCTATCAGATGACCCATTTGAGAGTGTATTCATCTTCAGGT 600  
DB 541 gatctcgtctcttctcatcagaatgcagccactttggagtgctatctatcttcaagt 600  
QY 601 CCCCTACAGATTTCAGTAATTAATCAGCAGAGATTAATTTGCACAAATACTGCCAAGAT 660  
DB 601 cccctacagatttcagtaaatcagcagagataagtttcacaaaactctgcagaagt 660  
QY 661 GCTTCGAGAGGCTTGTCTCCGATGCGATATGATTTGATTCAGAAATGCTTATGT 720  
DB 661 gcttcgagagcttgtctccgatttcgcgatgagatattgataatgacaagtgctatgt 720  
QY 721 CTATTTGAGAAATGAGCATTGACTACTCTATATAGTAAAGGAGGCGCTGTCTGTGT 780  
DB 721 ctatttcgaaaatagagcttgcactgcctatagtgaaaggagggctgtcgtgt 780  
QY 781 CTTCACACTTCAGGAAGTAGTACTCCAGTTCATTTGTGACTTCTGTGCAATAACAG 840  
DB 781 ctcccaacttcaggaagtagtactccagttccctatgcttctctgcacataacag 840  
QY 841 TTAGTCTTTAAAGAAACCTTCATTAATGCTGGCGAGCCATTTATGTAGAAACTT 900  
DB 841 ttagtcttttaaaagaaaccttcatttaagctggcgagccattatgctaggaact 900  
QY 901 AGCATCTCTCAGAGAGTCTCTACTTATTTATCAATTAATATCATATGCAATTCGCA 960  
DB 901 agcatctctcagagagctctactctatattataataataatcatatgcaattcgca 960  
QY 961 AATTATAGTGGACCTTATGCTTATGATGAGGAGATCAGTTTATCAGCAGAGAA 1020  
DB 961 aattatagtgagctatgtcattgacatgacatgagggagatcagttatcaagagaa 1020  
QY 961 AATTATAGTGGACCTTATGCTTATGATGAGGAGATCAGTTTATCAGCAGAGAA 1020  
DB 961 aattatagtgagctatgtcattgacatgacatgagggagatcagttatcaagagaa 1020  
QY 1021 GGAACATTTATCTCCAGGAACCGAGACGCTTACCGTTTGAATGGACATCCATCT 1080  
DB 1021 ggaacatttatctccaggaaccgagacgcttaccgcttcttgaatggacatccatct 1080  
QY 1081 TTACAAAATGCTAAATTCCTGAAATTAACAGCGAGAAATGATGCTATAGAAATTTAT 1140  
DB 1081 ttacaaaatgctaaatctccgaattacagcgagaaatgactctatagatattat 1140  
QY 1141 GATCTTATTTCTCTGAGAGATGGGTCTACCCATTTGAATTCACGAGATTCCTAAA 1200  
DB 1141 gatcttatttctctgagagatgggtctacccatttgaatttcacgagatctctaaa 1200  
QY 1201 AATAAGAGTACACAGAGACATCTCTTCTGAGAAAGAGTCTAGCAACGATCTCT 1260  
DB 1201 aataagagtacacagagacatctcttctgagaaagagcttagcaaacgatctct 1260  
QY 1261 AGGATTTTAAATCTCAATCCCTCAGAAACGTCAGCTGTCTCAGAGATCTTATGTT 1320  
DB 1261 aggatttttaaattctcaatccctcagaaacgctcgtctcagagatcttagttat 1320  
QY 1321 AAAGAGGGGCGCAAGTGCACAGTTTCAAAATTCACGACAGTCTCAGAGATCGATTAGT 1380  
DB 1321 aaagaggggcgcaagtgcacagtttcaaaatttcacgacagtctcagagatcgattagtt 1380  
QY 1381 TTAGATTTAGAACCAACTGATAGCTCTTAAGGAAGACATTGCCATACAGGCTCGCG 1440  
DB 1381 ttagattttagaaccaactgatagctcttaaggaagacattgccatcacaggctcgcg 1440  
QY 1441 ATGATATATAGATGCTTAAAGCTCATCTCAACAGCAGCTGTTATTAAGCAAAACACCG 1500  
DB 1441 atgatatatagatgctttaaagctcatctcaacagcagctgttataagaaacacgcg 1500  
QY 1501 AATAAAGAGTATCCGTCGACGACCTATATGAACTTATCCGCTACTGCAATGCTAT 1560  
DB 1501 aataaagagtatccgtcgcacgactatataagaacttaccgctactgcaaatgcttat 1560  
QY 1561 GAAGATCTCAGATGAGAAATTCACAGACGCTTCCCTGTGCTCTTATAGAGCTTGAGCC 1620  
DB 1561 gaagatctcagatgagaaattcacagagcttccctgtcctctttagagccttgagcc 1620



QY 1621 GGGGAGTAGTGTGACTGTAAGTCTGAGAGATTTCCTACCGGTAGTCCCATATGTTT 1680  
 |||||||  
 Db 1621 ggggagtagtgtgactgtgactgtagagatttcctaccggtgaagtcaccatagttc 1680  
 |||||||  
 QY 1681 CAAGGCAATTGGAAATTTAGCTTGGACAGAACTGGAAACAAAGTTGGAGATTCTTGG 1740  
 |||||||  
 Db 1681 caaggcaattggaaatttagcttggacagaaactggaaacaagtggagaattcttcg 1740  
 |||||||  
 QY 1741 GATTAATAATTTTAAGCTTACGCTGAGAAAAGAGAAATTTAGTTCTTAATATCTTG 1800  
 |||||||  
 Db 1741 gataaataaatttaagccttagaccgaaagaattagttccatattctg 1800  
 |||||||  
 QY 1801 TGGGGATGATGCTTAATATGTCAGATCCTTAATGAGGTCCAGAGCCATGATCGAGC 1860  
 |||||||  
 Db 1801 tggggatgactgttaattgacagatcccttaaggttcaagagaccatgacgagc 1860  
 |||||||  
 QY 1861 TTAACAGACAGATCGAGGGCTGTGGAGTGAATGGAATTTTCCATGATATGCC 1920  
 |||||||  
 Db 1861 ttacagacagatcgagggtgtgagatcgaaattgggaatttcccatgatactgccc 1920  
 |||||||  
 QY 1921 TCCGAAACAAATATTAAGTACCGTATACAGCGGTGATATGTTCTATCTGTAATAT 1980  
 |||||||  
 Db 1921 tccgaaacataataaagtaacggtacataacagcgtgagatagttctcatctgtaaat 1980  
 |||||||  
 QY 1981 GAGATCACACCTTAAGCACTATCTTCATGAGTGCATTTCCCACTCTTAGTAGACAG 2040  
 |||||||  
 Db 1981 gagatcacaccttaagcactactactgtagtgcatttcccaactcttagtagagaaag 2040  
 |||||||  
 QY 2041 GACTATGCGGTTTCCAAACAGAAATGATGATTTAGGATGATCTCTATCAATAT 2100  
 |||||||  
 Db 2041 gactatgcggtttccaaacagaaatgatatgatttagatgtagtactctcatat 2100  
 |||||||  
 QY 2101 ACAACCTCCCTAGGAGATATTTCCCTATGCTTCGCGTAACTTAATGTAACGTGGG 2160  
 |||||||  
 Db 2101 acaacctcccttaggagatatttccgttactgcttgcgttaaccttaagtaaacgctgc 2160  
 |||||||  
 QY 2161 ATTCTCTCAAGAGGTTCTTCAAAATCCTATATGATTTTCTTTTCTGCTGCTTAT 2220  
 |||||||  
 Db 2161 attctctcaagagggttcttcaaaaacctctttagatttcttcttcttcttcttctt 2220  
 |||||||  
 QY 2221 GGTATGCCCAACCAATGATGAAAAACAGACTACGCAATTTCCCTATGTAAGGAAAAACAGC 2280  
 |||||||  
 Db 2221 ggtatgcccaacaaatgatatgaaaaagactacgcaatttccctatgtagaataacagc 2280  
 |||||||  
 QY 2281 TGGGAAACAATTTGCTGGCTATAGAGTGGGAGAGACATGCTCTATGTTGATGAG 2340  
 |||||||  
 Db 2281 tgggaaacaatttggctgataagtgaggagagacatgctctattgtagtattgag 2340  
 |||||||  
 QY 2341 AACGAGAGACTTTTCCAAAGTGCATCCATTTATGAACATACAAATAGTTATGTTAT 2400  
 |||||||  
 Db 2341 aacgagagacttttccaaagtgcatacccatltagaaactaacatagtttagtctat 2400  
 |||||||  
 QY 2401 CAGGAGACTTTCAAGACGACTGCAGATGCCGTAGATTAGTAATGGAGATTAAACA 2460  
 |||||||  
 Db 2401 caggagactttcaagacgactgcagatgcccgtatgattagtaagagatttaaca 2460  
 |||||||  
 QY 2461 TCGATTTCTGCTACTCTAGGCAATCGTTGAGAGAGTGCACATTTCTCAGAGATGATC 2520  
 |||||||  
 Db 2461 tcgatttctgactcttaggcaatcgcttggagaagcttgacacttccagatgatactc 2520  
 |||||||  
 QY 2521 TATGACTTTGTTTCTCTATATCTGATATTTTCCGTAAAGATGCCCTATGTAAGCT 2580  
 |||||||  
 Db 2521 tatgactttgttctctatattccctgatatatttccgtaagatccctcatgtagaact 2580  
 |||||||  
 QY 2581 GCTCTGCTGATTAAGCGAGACTCCTGCTGTTGCCGACACACATGATCAAGACATGCT 2640  
 |||||||  
 Db 2581 gctctgctgatttaagcagactcctgcttgcgcgacacacgataagacatgct 2640  
 |||||||  
 QY 2641 TTTGTAGGAGTGAAGGCTGATGATGTAATGAGTCTTATGTCGA 2700  
 |||||||  
 Db 2641 ttgttaggagtgaaagctgatacttaacgactatactgagccttattgctga 2700  
 |||||||  
 QY 2701 GGAAGTATAGAAATGCCCCCATGCTAGGAATTAATTAACCTGGAACCAATTT 2760  
 |||||||

Db 2701 ggaagtagaagtcgcccccatgtagagataataataaactgtggaagcaattt 2760  
 |||||||  
 QY 2761 CGTTTATGAAGTTTCCATGCTGTGTGCTCCGATCTTAATTAATCTGAGCT 2820  
 |||||||  
 Db 2761 cgttttagaagtttccatgctgtgtgttccgacttaactaataactcgtgact 2820  
 |||||||  
 QY 2821 ATGATCATAGGCAATTTGGGTTTCTCGAACTGTGTGGAGATTAACGATTTTATGCA 2880  
 |||||||  
 Db 2821 atgatacataggaactgtgttctcgaaactgtgtggaataacgacattatagca 2880  
 |||||||  
 QY 2881 TAAAGCAATCTGATACACCTCAGCCCTAGAGACATTTCTTAGGGGTTCTTATTTGT 2940  
 |||||||  
 Db 2881 taagcaatctgatacaactcagccctagagacatctttaggggttcttattgt 2940  
 |||||||  
 QY 2941 CTAACCTTCGATTTATATCGAAATCCTTAAGTTCCTGTTGCTCTCCAGAGACT 3000  
 |||||||  
 Db 2941 ctaaccttcgattttagcgaatcccttactgcttctgttctgttctgcgcgagagt 3000  
 |||||||  
 QY 3001 TCTTAACGAATCATAGGATTCAGGGTCTGTTCCCTGAGTCTTGGCA 3052  
 |||||||  
 Db 3001 tcttaacgaatcataggaattcaggggtctgttcttcttgcgccttgca 3052  
 |||||||

## RESULT 2

AAK91990  
ID AAK91990 standard; DNA: 1230025 BP.

AC AAK91990;  
 DT 13-SEP-1999 (first entry)

DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; ss.  
 OS Chlamydia pneumoniae.

PN WO9927105-A2.  
 PD 03-JUN-1999.  
 PF 20-NOV-1998; 98MO-IB01890.  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.

XX (GEST ) GENSET.

XX Griffiths R;

XX WPI; 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

PS Claim 1; Page 291-611; 1912pp; English.

CC The present sequence represents the complete genome of Chlamydia  
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes  
 CC respiratory disease such as pneumonia and bronchitis and is thought  
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,  
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
 CC encoded by the open reading frames of the C. pneumoniae genome (see  
 CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.  
 CC Vectors containing C. pneumoniae nucleotide sequences can also be  
 CC used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae.

Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match	99.5%	Score 3037.8	DB 20	Length 1230025
Best Local Similarity	99.7%	Pred. No. 0		
Matches 3042	Conservative 0	Mismatches 10	Indels 0	Gaps 0
1	ATGCGATTTTGGCTCGCGGATTTTCTCTAGTTTCTTTTAACTTGCTCCAGCTTC	60		
14484	atgagatttcgctcgcggatllccctcaglltllcttcaaatgctccagcttc	14543		
61	GACACTTCTTTTAGTGTCTCTACGATTTCTTTTACCCCGAAGATAGTTTATGAGAT	120		
14544	gacacttctttagtgcgtactacgattctttaaaccocgaagatagtttcatgaggt	14603		
121	AGTCAGATGCAAAACGTTTATATATGTTCAAGCTGGGAGTGTATAGCTTACTGCT	180		
14604	agtcagaatgcaagaacgttctataatgtccaagctgggagtgctatagcttactggt	14663		
181	GATGTCTCAATTTCTACAGTGTATACCTTGATTAATTAAGCCCTGCTCAATGTGACC	240		
14664	gatgtctcaaatctcacaagctgaatcctcgatttaaaagaaccgtcttcacgtgacc	14723		
241	TCAGGAAGTGTACCTTTCGACAGAAATCATCATGGGTTTATTTTAAATATTTTCCCTCA	300		
14724	tcaggaagtgtagcgttcgcagaagaatcatcctbgygtnatatttcaataatcttccca	14783		
301	GGACTTACAAAGGAAGGGGCTGTACTTTGTGTCACAGATCTCTCAAGCAACGGACGTTT	360		
14784	ggaactacaagaagagngctgtacttgtgtccaagatcctcacaagcagcacgcttll	14843		
361	TCTGGGTTCTCCAGCCTCTCTTTTATTCAGACCCCGAGATATTAAAGAACAGGATGT	420		
14844	tcctgggtctccaagcctctctttaaacaagccccggagatttaaagaagaagagat	14903		
421	CTCTATTCAAAAAATGACACTTATGCTTTAAACAATTAATGTAGTCGTTTGAACAAAAC	480		
14904	ctctattcaaaaaatgcaactatgcttcttaaacaattatgtcgttcttgaacaaaaac	14963		
481	CAAGTAAAGACTTAAAGCGGAGCTATTAGTGGGGCGAATGTATTATAGTAGGCACTAC	540		
14964	caaaagtaagactaaaggcgagctattagtgggcgaaatgtactatagtaagcaactac	15023		
541	GATTCGCTCTCTTTATCAGATACGACCACTTTTGGAGGTCATTCATCTTCAGGT	600		
15024	gattcgcctctcttcatatcaagaatgcaagcacttlltgagggtcatccatcttccagtt	15083		
601	CCCCACAGATTGCAGTAAATCAGCGCAGAGATAAAGATTGTGCAAAAATATCTGCCAAGAT	660		
15084	ccccacagattgcaggtaatatcaggcagagataaagatttgcacaaaatcgtccagaagt	15143		
661	GGTTCGAGGGGGCTTGTACTCGCGATGGGTGATTTGATTCAGAAATGCTTATGTT	720		
15144	ggttctggaagggtcttgaactccgaatggtgatattgatatgacagaatgcttatglt	15203		
721	CTATTTCGAAAAAGAGGCACTGACTCGTGTATAGGAAGGAGGGGCTGTCGTGT	780		
15204	ctatttcgaaaaagagagcatgactgctcgtctaaagaaagggggggcgtcgtgtgt	15263		
781	CTTCCACATTTCAGGAAGTAGTACTCCAACTCTTATTTGACTTTCTCGACAAATTAACAG	840		
15264	cttccacatttcaggaagtagtaactccagttcctactatgtgacttctctcgacaataacag	15323		
841	TTTACTTTTGAAGAACAATTCCTAATATGGGTGGGAGGCCATTTATCTTAGAAACTT	900		
15324	tttagcttcttgaagaacaatcatcctaataatggtgcygagccaattatgtctaggaactt	15383		
901	AGCATCTCTTCAGGAGGTCCTACTATTTATCAATATATATCATATGCAAAATTCGCA	960		
15384	agcatctcttcaggaaggtctactctatctatccaataataatacatatgcaaatctgcaa	15443		
961	AATTTAGGTGGAGTATTCGCAATTGATCTGGAGGGGAGATCAGTTTTCAGCAAGAGAA	1020		
15444	aatttaagtgtagcatatgtccatltgatacttggaaagggagatcagttatcagcagaagaa	15503		
1021	GGACAAATTTACATTTCCAAGAAACCGGACGAGCTTATCCGTTTTCATTAAGGCAATCATCTT	1080		

D	15504	ggaacaattacac	ttccaaagaaac	cggaagc	ttaacg	ttcttgaa	tggaac	ttc	15563
O	1081	TTTACAAAAATG	CTAAATTCCTG	AAATTTACAGG	GAGAAATG	GTGCTTAA	AAATTTAT		1140
D	15564	ttacaaaatgct	caaatcccttg	aattccag	gcgggaag	aatgga	tactctat	agaattttat	15623
O	1141	GATCTATTACT	TTCTGTAAGC	AGATGGGTC	TACCCAA	TTGTAATAT	CAACGGAG	ATCTTAA	1200
D	15624	gattcctattact	ctcttgaag	cagatgg	gtctacc	ccaattga	tatacaag	gagatcc	15683
O	1201	AATTAAGAT	ATACACAGG	ACCATACT	CTTTTCTG	GAGAAAAG	ATGCTTAG	CAACCATCT	1260
D	15684	aaataaagata	cacagggac	taactctt	cttcttg	gaaaaag	atctatg	caacg	15743
O	1261	AGGATTTTAA	ATCTACATCC	CTCGAAGC	GCACACCT	CTGCAAG	AGATTTAGT	TTAT	1320
D	15744	agggatttcaat	ctcacaatcc	ctccaga	gcgtcaac	ctgtctcg	cgga	tacttga	15803
O	1321	AAAGAGGGG	CGGAAGTC	ACAGTTTCA	AAATTCACG	CACTCTC	CAAGATG	CAATTAGT	1380
D	15804	aaagagggg	ccgaagtc	acagtttca	aaattc	acgcag	ttcc	caagatgc	15863
O	1381	TTTAGATTT	AGGAACAA	ACTGATAG	CTTTAA	AGGAAG	ACATTGCC	ATCAGACG	1440
D	15864	ttagatttag	gaacaaact	gtatag	ctctta	agaaag	atctg	ccatcaag	15923
O	1441	ATAGATAT	GTAGCTTAA	GCATCCTCA	ACAGACG	CTGTATTAA	CAACCAAC	ACCGCA	1500
D	15924	atagatataga	cttagctta	agcctca	ctccaca	gcgcgt	tattat	aaagcaaac	15983
O	1501	AATTAACAG	ATATCCGTG	ACGGACTT	ATAGAACT	TATCTCG	CTACTG	GCATCGCTAT	1560
D	15984	aataaacaga	tataccgt	gcgcga	ctctataga	acttact	ctgc	tacitg	16043
O	1561	GAAAGATCT	CAGATGAG	AAATTCAC	AGACGTTCC	CTCTCTCT	CTTTTAG	AGCCTG	1620
D	16044	gaagatctca	gaatbga	aaattca	gaagac	gttccct	ctgtct	ctcttag	16103
O	1621	GGGGGAG	CTACAGT	TAACTG	CTGAGATTT	CTTACCG	CTAAGTCC	CAATATG	1680
D	16104	gggggaag	tgacgt	ctaac	ctgc	tggaag	ttctcc	taacg	16163
O	1681	CAAGCAAT	TGGAAAT	TAGCTTG	AGACAG	ACATG	GAACAAAG	TTTGAAAT	1740
D	16164	caagc	caattg	aaattag	cttgc	agcaga	acttga	aaacaag	16223
O	1741	GATTAAT	TAAATTA	TAAAGCT	TGACCTG	AAAAAG	AAAGAAAT	TTAGTCT	1800
D	16224	gataaata	ataattca	taagct	cttag	accttga	aaagaag	aaattag	16283
O	1801	TGGGGAA	TGCTGTAA	ATTCATG	ATCTTA	TGCAAG	GTTCAG	AGAACCCAT	1860
D	16284	tgggggaat	gctgtat	ga	tgtca	gcttca	atgc	aggttca	16343
O	1861	TTTACAG	ACAGATG	AGGGCTG	TGTGATCG	ATGTGA	ATTTGG	GAATTTCT	1920
D	16344	ttacagac	agatc	gsggggt	gtg	ggatcg	atgtg	gaattcg	16403
O	1921	TTCGAA	AGACAA	TATTAAG	GTACCG	CTATAAC	AGCGTG	ATATGTTCT	1980
D	16404	tccgaaga	aaataata	aggtac	cgctaca	taacag	cgg	tgatata	16463
O	1981	GAGATC	ACACCC	TAAACAC	ATATCT	TCGTG	ATTTGCC	AACCTTTA	2040
D	16464	gagatcca	accctaa	gcacata	ctctcg	atgtg	aatltt	cccaact	16523
O	2041	GACTAT	TGCGGT	TTTCC	AAACA	CAATAC	AGAAATG	ATTTAG	2100
D	16524	gactatg	cggltt	ccaaca	agaa	tacaga	atgtat	tttag	16583
O	2101	ACAAC	CTCC	TAGGA	ATATTTT	CCGTTAT	GTGCG	CTAAAC	2160

```

Db 16584 acaacctccctagagatatcttcgcttagcttcgtaaccctaagtcaacgctgag 16643
QY 2161 ATTCTCTCAAGAGGTTTCTTCAAAATCCTTATGATTTTATTTTGTGTGCTTAT 2220
Db 16644 attctcccaagaaggttcttcaaaatccctctatgatttcttcttctgtgctatc 16703
QY 2221 GGTATGCCACCAATATATGAAGAGACTAGCAAAATTTCCCTATGTAAGAAACAGC 2280
Db 16704 ggtcattcccaaccatgatatagaacaagactaagcaatttccctatggtgaaacagc 16763
QY 2281 TGGAGAAACAATTGTTGGCTATAGAGTGGCGAGAGAGATGCTCTATTTGATTTGAG 2340
Db 16764 tggagaacaacttctgtggctatagagtgagggagagcattgctctatgattgag 16823
QY 2341 AAGGAAAGACTTTTCCAAAGTGGCATCCATTATGAACATAAATTAAGTTATGCTTAT 2400
Db 16824 aaggaagaacttccaaagtgcattccattatgaacaataatgatttctat 16883
QY 2401 CAGGAGATTTTCAAAAGAGAGAGAGATGCGCGTATGTAATGAGAGATTGAAACA 2460
Db 16884 caggaggaattccaaagaagactgagatgcccagattgataatggagattca 16943
QY 2461 TCGATTCTGTACCTCTAGGACATACGCTTGGAGAACCTGGCACTTCTCAGATGATCT 2520
Db 16944 tcgattctgtacctctagagcatalagcattgagaagctgcaacttctcagagtgatc 17003
QY 2521 TATGACTTTAGTTTCTCTATATATCTGATATTTTCCGTAAGATCCCTCATGTGAAGCT 2580
Db 17004 tatgactttagttctctctatctctgataatttccgtaagatccctcatgtgaagct 17063
QY 2581 GCTCTGTGATTTAGCGAGAGACTCTGCTGTTGTCGGAGCAACAGATACATCAATGCT 2640
Db 17064 gccctgtgatttagggagaaactctgtctgtccgcaagcaacatcaaaacatgct 17123
QY 2641 TTTGTAGGAGAGTGAACGGGTCGATACCTTTTAACTATATCTGATCTTATGCTGA 2700
Db 17124 ttgttaggagtggaacggtgcgtgatacacttaacgactatcagctcttattgca 17183
QY 2701 GGAAGATATGAAATGCCGCCCATGCTAGGAATTAATTAATTAACGTGGAACAAATTT 2760
Db 17184 ggaagatatgaatgcgcgcccatgctaggaatataataaactgtgaaagcaaat 17243
QY 2761 CGTTTTAGAAGGTTTCCATGCTGCTGTGTTCCGATCTTAACTATTAATCTGGAGCT 2820
Db 17244 cgttttagaaggttccattgcctgtgtgtgtcggagcttaactataatcctgagct 17303
QY 2821 AAGGATCATAGGCAATGCGTTCTCGAATTTGTGAGAGAAATTAACGACATTTTATGCA 2880
Db 17304 atgatacatagagcattggtctctcgaactgtgtgagaataagacatttataatgca 17363
QY 2881 TAAAGGAATATCTGATACACCTCAGGCCCTAGAGACATTTTAAAGGGTTCTTATTGT 2940
Db 17364 taaggaataactgtaaccctcagccccagagacattcttaaggttcttattgt 17423
QY 2941 CTAACCTTGTATTTATGAGAAATCTTTACCTTCTGTTGTTGTTGTTCTCCGAGAGAT 3000
Db 17424 ctaaaccttgatattatataagagaatccttaagttctgtgtgtcctgcgagagagt 17483
QY 3001 TCTCTAACAAATCATAGGATTCAGAGGTTCTGTTCTTGTAGTCCCTTGGCA 3052
Db 17484 tctctaagaatcatagaggttccaggttctgtcttcttgagtccttggca 17535

```

RESULT 3  
ID AAA50035 standard; DNA; 2950 BP.

XX AC AAA50035;  
XX  
DT 10-OCT-2000 (first entry)  
XX  
DE DNA encoding Chlamydia pneumoniae antigen CPN100877 RY-61.  
XX

```

KM CPN100877 RY-61; antigen; infection; diagnosis; therapy; vaccine;
KW outer membrane protein; ds.
XX Chlamydia pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 101..2869
FT /tag= a
XX
PN MO200039158-A1.
XX
PD 06-JUL-2000.
XX
PF 23-DEC-1999; 99MO-CA01230.
XX
PR 23-DEC-1998; 98US-0113280.
PR 23-DEC-1998; 98US-0113281.
PR 23-DEC-1998; 98US-0113282.
PR 23-DEC-1998; 98US-0113283.
PR 23-DEC-1998; 98US-0113284.
PR 23-DEC-1998; 98US-0113285.
PR 23-DEC-1998; 98US-0113385.
PR 28-DEC-1998; 98US-0114050.
PR 28-DEC-1998; 98US-0114056.
PR 28-DEC-1998; 98US-0114057.
PR 28-DEC-1998; 98US-0114058.
PR 28-DEC-1998; 98US-0114059.
PR 28-DEC-1998; 98US-0114061.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Murtin AD, Oomen RP, Wang J;
XX
DR WPI: 2000-452369/39.
XX
DR P-PSDB: AAY95548.
XX
PT Novel Chlamydia polynucleotides and polypeptides useful for diagnosis,
PT prevention and treatment of Chlamydia infection in mammals -
XX
PS Claim 2(a); Page 65-69; 215pp; English.
XX
PS
XX
CC The present sequence is that of Chlamydia pneumoniae genomic DNA
CC including an open reading frame that codes for CPN100877 RY-61 (see
CC AAY95548), a putative 98 kDa outer membrane protein. It is an example
CC of C. pneumoniae polynucleotide molecules of the invention (see
CC AAY50030-42) that encode antigenic polypeptides (see AAY9543-55) useful
CC in the diagnosis, treatment and prevention of Chlamydia infection.
CC The polynucleotides can be utilized: in the recombinant production
CC of Chlamydia antigens using transformed unicellular host cells; in
CC vaccines or live vaccine vectors; in naked form or formulated with
CC a delivery vehicle for therapy and prophylaxis of Chlamydia
CC infection; and as probes or primers for diagnosis of Chlamydia
CC infection.
XX
SQ Sequence 2950 BP; 854 A; 576 C; 626 G; 894 T; 0 other;

```

Query Match 93.2%; Score 2843.6; DB 21; Length 2950;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2846; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 AAGCGATTTTTCGCTCGCGGATTTCTCTAGTTTTCCTTAACATGTCGTCAGTCTTC 60
Db 101 atggcatttctgcctcgggattcctcctagtttcttcttaccatgctcctagctctc 160
QY 61 GACACTTCTTGTAGTCTCTACGATTTCTTAAACCCAGAGATAGTTTTCATGGAGAT 120
Db 161 gacacttcttggtagtctactacgatttctttaaaccagagatagtttctcatggagat 220
QY 121 AGTCAATGACAGACGTTCTTAATGTTCAAGCTGGGGAGTCTTAAGCTTACTGCT 180
Db 221 agtcagaatgcagaagcttctataatgttcaagctggtgagatgtcatagccttactgct 280

```

QY	181	GAGGTCGAATATCTAACGTCGATTAATCTGCAATTAAATAAAGCCGCTTCAATGTCAC	240
Db	281	gatgctcaataatactaacgctcogaataacctgcgaataaataaagccgcttcaaigtgcac	340
QY	241	TCAGCAAGCTGTGAAGTTTCGACGAGAAATCATCATGAGGTATATTTTTAATAATATTTCTCA	300
Db	341	tcaggaagtgycagcttcgcaggaatacatcaygtttatatttaataataattcccca	400
QY	301	GGAGCTTCAAAAGAAAGGGGCTGTACTTTGTGGCAAGATCCCAAGCAAGCGACAGTTT	360
Db	401	ggaactacaagaagagggcgctgaacttcttgccaagatccccaagcaacgycagtttc	460
QY	361	TCTGGGTCCTCAGCGCTCTCTTTTATCAAGCCCGGAGATATTTAAAGAACGAGATG	420
Db	461	tcgtggtctccacgctcctctttatttcagagcccgagatataaagaacaggaatgt	520
QY	421	CTCTATTTAAAAAATGCACTTATGCTTTTAACAAATATATGTAGTCGTTTGAACAAAC	480
Db	521	ctctattcaaaaaatgaactatgctctcttaaaacaattatgtagtgcgttttgaacaaac	580
QY	481	CAAGTATGACTTAAAGCGCGAGCATATAGGGGGCAAGTTACTATATATAGCAACATAC	540
Db	581	caagatgaactaaagcgagctatagtggggcgaaigttaactatagtagaacctaac	640
QY	541	GATTCCGTCCTTTCTATCAGAAATGCAAGCCACTTTTGGAGGTGCTATCCATTCTTCAAGT	600
Db	641	gattccgctctcttctctacgaatgcagccactttggagtgctatccatcttcaggt	700
QY	601	CCCCACAGATTGCGAATAATCAAGGACAAATTAAGATTAATTTGCACAAATATCTGCCAAGAT	660
Db	701	ccccacagatctgagtaaaatcaggcaggaataagcttcgcacaaataactgcgaagat	760
QY	661	GGTTCCTGGAGGGCTTTGTACTCCGATGCTGATATTTGATATTTGATCAGAACTCTTATGT	720
Db	761	ggttcctggaggggcttcttgtaactccgatgctgatatctgatattgatcagaatgcttatgtt	820
QY	721	CTATTTGCGAAATGAGGCAATGACTACTGCTATAGTATTAAGGAGGGGCTGCTGTGT	780
Db	821	ctatttcgagaagaaigtgagatctgactactgcatagttaaaggagggcgctgcctgtgt	880
QY	781	CTTCCCACTTACGAAATATATCTCCAGTTCTCATTTGTACTTTCTGCACAAATTAACAG	840
Db	881	cttcacaacttcaggaagtagtaactccaagttcccatgtgtaacttctcgcacaataaag	940
QY	841	TTTAGCTTTGAAGAAACCATTCATATAGGGGTGGCGGACCAATTTATGCTAGAACTT	900
Db	941	tttagtcttgaagaagaacccatccataatgygtggcgagccattatgctagaacct	1000
QY	901	AGCATTCCTTAGAGAGTCTCTACTCTATTTATCAATAATATATTCATATGSCAAATTCGAA	960
Db	1001	agcatctcttcaggaagctccactcattatcaataataatacaatcagaacttcgaa	1060
QY	961	AATTTAGGTGAGCTATTTGCATTTGATCTGAGGGGCAATCAGTTTATCAGACAGAAA	1020
Db	1061	aatttagtgtagtattatgcatgtatgctagtcgggggagatcagttatcagcagagaa	1120
QY	1021	GGAAACATTTACATTTCCAAAGAAACCGAGAGAGCTTACGCTTTTGAATGGCATCATCTT	1080
Db	1121	ggaaacaatcatcatccaaggaanaacgagacgagttacgcttttgaatgcatccatct	1180
QY	1081	TTTCAAAATGCTAAATCCCGAATTAACAGGCGAGAAATGATGCTATAGAAATTTAT	1140
Db	1181	tttaaaaaatgctaattcccgaaataaagcgagagaaatgtagtactctatagaattctat	1240
QY	1141	GATCTTATTTACTTCTGAAGCAGATGGGCTTACCCAATGTAATATCAAGGAGATCTTAA	1200
Db	1241	gatctcttactctcgaagcagatggtgtctaccaatltgaatatacaagagatctcaaa	1300
QY	1201	AATTAAGATGACACAGGACCATATCTTTTTCGGAGAAATAAGTCTAGCAAAAGCATCT	1260
Db	1301	ataaagagtagtaacaaggaactaactccttcttcggagaaagagcttagcaacgaacct	1360
QY	1261	AGGATTTTAAATCTCAAAACCTTCAGAACGTCAACCTGTCTGCGAGGATCTTATGTTT	1320

Db	1361	agggatttaattcaatccctccgaagaactgcttcgcggataacttgatttt	1420
Oy	1321	AAAGAGGGGCGCAAGTCACAGTTTCAAAATTCAGCCAGCTTCACAGATCGCATTTAGTT	1380
Db	1421	aaagagggggcgcgaagtcacagatctcaaaatcaacgcagctcccgagatcgcatttagtt	1480
Oy	1381	TTAGATTAGGAACCAACGATAGAGCTCTTAAGGAAGACATTGCCATTCACAGCCCTCGG	1440
Db	1481	ttagatttaggaaccaaacgcgaatagctcccaaggaagaactgcacacagcctcgcg	1540
Oy	1441	ATAGATATAGATAGCTTAAAGCTCATCTCAACGACGAGCTTTATTAAGCAAAACCGCA	1500
Db	1541	atagatataagtagcttaagctcatccctcaacgagcgcgtattataagcaaacgcgca	1600
Oy	1501	AATAAACGATATCCCGACGGACTGTATGAACTTATCTCCGCTACACGCAATGCCAT	1560
Db	1601	aataaacgaatataccgcgcgaagactcaatgaactatactgcctacgcgcgcgcgcgc	1660
Oy	1561	GAAGATCTCAGAAATGAAATTTACAGAGCTTCCTCTCTCTCTCTTTAGAGCTTGAGCC	1620
Db	1661	gaagatctcgaataggaagaattcaagaagcttcccctcgtctctcttttagagcctggagc	1720
Oy	1621	GGGGTAGTGTGACTGTAACTGCTGGAGATTTCTTACCCGTAAGTCCCATATGCTTTT	1680
Db	1721	ggggtagtgcgactgtgaactgccttggaagatttctccacgcgtagtctccattatgcttt	1780
Oy	1681	CAAGCAATTGAAATTAAGCTGTGACAGCAACGGAACAAAGTTGGAAATTTCTTCGG	1740
Db	1781	caagccaattggaaatagcttggacaggaagaccgcggaacaaagttgggaattctcttcgg	1840
Oy	1741	GATAAATTAATTATTAAGCTTAAGCTGAACCTGAAAAAGAAATTTAGTTCATATCTTG	1800
Db	1841	gataaaatlaattlaagccttagacctgaaagaaagaaatlltagtctcctaattcttg	1900
Oy	1801	TGGGGGAATGCTGTAATGTGACATCCCTTAATGACAGTTCAAGAGCCCATCATCGAGC	1860
Db	1901	tgggggaatgcgtgaagtgcgaatccctaaagcgaagttccaagagcccaagcagcgcgc	1960
Oy	1861	TTACAGACAGATCCAGGCGCTGTGGATCGATCGATGGAATTTGGAAATTTCTTCATGTATCGCC	1920
Db	1961	ttacagacagatcgcgagggcgtgtagatcgatggaaatttcttccaagtatctgc	2020
Oy	1921	TCCGAAGCAATATTAAGGTACCGTCAATACAGCGGTGGATATGTTCTATCTGTAATAT	1980
Db	2021	tccgaagaacaataaagtatccgtcatacaagggggtgaatagtctatctgtaaat	2080
Oy	1981	GAGATCACACTTAAGCCTATACCTATCTGAGGGCAATTTTCCCACTCCTTTAGTGAACAAG	2040
Db	2081	gagatcacacctaaagcactactactcgaagcatttctccaactcttcagtagagacaag	2140
Oy	2041	GACTATGCGGTTTCCAAACAAGATACAGATATGATTTAGATGCTATCTCATATAT	2100
Db	2141	gactatgctggtttccaacaacagatcacgaatgatttagtagctgatactctatacat	2200
Oy	2101	ACAACGTCCTTAGGGATATTTTCCGTTATGCTGCGGTAACCCATATGTTAAAGTCGGG	2160
Db	2201	acaacgtccctcgaaggaatatcttcggtatgcgttcgcgtaaccctaaagtaaacgcgcgg	2260
Oy	2161	ATTCTCTCAAGAGGTTTCTTCAAAATCTCTTATGATTTTTCATTTTGTGTCCTAT	2220
Db	2261	attctctcaagaaggtttcttcaaatccctctatgattttcatatttttgggcttat	2320
Oy	2221	GGTATGCGCACATGATATGAAACAAGATACGCAAAATTTCCCTATGCTGAAAAACGC	2280
Db	2321	ggtatgcccccaatgatactgaaaacaagactcgcgaatcttcccataggctgaaaaaagc	2380
Oy	2281	TGGAAGAAACATTTGTTGGGCTATAGAGTCCGAGGAGAGCATCCTCTATTTGGTATTATAG	2340
Db	2381	tgggaagaaacattgttgggctataagagctgcggagggagacaagcctctatggatttggag	2440
Oy	2341	AACGGAAGACTTTTCCAAGGTGCGCATCCATTATGAACATCAATATGATTATGCTTAT	2400

Db 2441 aacggaagactttccaagtgccatccattatgaactaactagttatgcttat 2500  
QY 2401 CAGGAGATTTCAAAGACACACTGCAGATGCGCGTAGATTATTAATGGAGTTTAA 2460  
Db 2501 catgagattccaagaacgactgcagatgcccgtatgattagtaatgaggtttaca 2560  
QY 2461 TCGATTCTGACTGACCTAGGATACGGCTTGAAGCTGGCAGCTTCTGCAGATGTACTC 2520  
Db 2561 tgcattctgtaacctcagagcaatagcttgagagaagcggagacttccagagatgtaactc 2620  
QY 2521 TATGACTTATGATTTCCTCTATATTCCTGATATTTTCCTGTAAGATCCCTCATGTGAAGCT 2580  
Db 2621 tatgactttagttctctctatattcccgatatttccgtaaggtccctcaatggaagct 2680  
QY 2581 GCTTGCTGATTAGCGAGACTCTGCGTGTTCGCGACAGACAGTATCAAGCATGCT 2640  
Db 2681 gctctggtgattagcagagactcctgctgttccgagcagacacagatcaagaacatgct 2740  
QY 2641 TTTGTAGGAGTGGAGGGGCGGTATCAGCTTAAAGCATATGAGCTTGTATGCGA 2700  
Db 2741 ttgtcagaggaatggaacgagcggtacgtaacacttaacgactaactgagctctatgctga 2800  
QY 2701 GGAAGTATGAGATGCCGCCCATGCTAGGAAATTAATATAAATGTAAGCAAAATTT 2760  
Db 2801 ggaagatagaaatgcccgcacccatgctaggaattataataaactgtggaagcaaatlc 2860  
QY 2761 CGTTTATGAAGGTTTCCATTGCTGTGTGTTCCGATCTTAACTATAATCTGAGACT 2820  
Db 2861 cgttttagaaggttccatctgctgtgtgttccgagatcttaactaataatcctgact 2920  
QY 2821 ATGATCATGAGGATTTGGTTTTCGAACT 2850  
Db 2921 atgatacacaagcattggttctcgaact 2950

RESULT 4  
AAK06829  
ID AAK06829 standard; DNA; 1200 BP.  
AC AAK06829;  
XX 26-APR-1999 (first entry)  
D7  
XX Chlamydia pneumoniae surface exposed protein Omp6 DNA.  
XX  
XX Omp6; outer membrane protein 6; surface exposed protein; antigen;  
KM Infection; diagnosis; vaccine; atherosclerosis; asthma; ss.  
XX  
OS Chlamydia pneumoniae.  
XX  
XX W0958953-A2.  
PN  
XX 30-DEC-1998.  
PD  
XX 19-JUN-1998; 98WO-DK00266.  
PF  
XX 23-JUN-1997; 97DK-0000744.  
PR  
XX (BIRK/) BIRKELUND S.  
PA (CHR1/) CHRISTIANSEN G.  
XX  
XX Birkelund S, Christiansen G, Knudsen K, Madsen A;  
PI Mygind P;  
XX  
XX WPI; 1999-105610/09.  
DR P-PSDB; AAW88430.  
XX  
XX Species-specific test for identifying mammals infected with  
PT Chlamydia pneumoniae - comprises detecting antibodies specific for  
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding  
XX these proteins  
PS Disclosure; Page 80-82; 115pp; English.

XX  
CC This DNA sequence encodes the novel surface exposed protein Omp6  
CC (see AAW88430) of Chlamydia pneumoniae, a human respiratory pathogen.  
CC It is described as a subsequence of a claimed nucleic acid fragment.  
CC (see AAK06818) encoding Omp6 (see AAW88419). The invention provides a  
CC new species specific test for identifying mammals (including  
CC humans) infected with C. pneumoniae. The test comprises detecting  
CC antibodies specific for surface exposed proteins Omp6-Omp6s (see  
CC AAW88417-28) or detecting nucleic acid fragments encoding them (see  
CC AAW88416-27), especially by PCR. The proteins are also used in the  
CC diagnosis of C. pneumoniae infection in mammals. The nucleic acids  
CC and proteins can also be used in the immunization of mammals, the  
CC nucleic acids being particularly useful as DNA vaccines for  
CC effecting in vivo expression of antigens. The vaccines may also  
CC prevent atherosclerosis and bronchial asthma, which are possibly  
CC associated with C. pneumoniae.  
XX  
SQ Sequence 1200 BP; 378 A; 243 C; 247 G; 332 T; 0 other;

Query Match 39.1%; Score 1192; DB 20; Length 1200;  
Best Local Similarly 99.6%; Pred. No. 0;  
Matches 1195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1192 GATCCTAAATAAAGGTACACAGGGACATCTCTTCTGAGAAAGATCTAGCA 1251  
Db 1 gatcctaataaaggtacacagggacatctcttctgagaaagatctagca 60  
QY 1252 AACGATCCTAGGATTTTAAATTCACATCCCTCAGAACGTAACCTGTGCGAGATAC 1311  
Db 61 aacgactcagggattttaaactcaaatccctcagaagcgaacgctcgtcagagatc 120  
QY 1312 TTACTTTTAAAGAGGGGGCCGAAGTCAACAGTTTCAAAATTCACGAGTCTCCAGATCG 1371  
Db 121 ttacttttaaagaggggccgaagtcaacagtttcaaaattcacgagctccagagatcg 180  
QY 1372 CATTTAGTTTATGATTTAGGAACCAACTGATAGCTCTTAAAGAAAGATGGCATGCA 1431  
Db 181 catttagttttagatttagaacaactgataagctctcctaagagaagatgcatcaca 240  
QY 1432 GGCTCCGATAGATATATGATAGTAAAGCTATCCTCAACAGCAGCTGTATTAAAGCA 1491  
Db 241 ggctccgataagatatagatagcttaagctcaccccaagaagcagctgtattaaagca 300  
QY 1492 AACACGCCAAATTAACAGATATCCGTGACGACTCTATAGACTTATTCGCTACTGCG 1551  
Db 301 aacacgcgaataaacaagatatccgtgacgactctataagactatctcgtactcgc 360  
QY 1552 AATGCTATGAAATCTCAGAAATGAGAAATTCACAGACGTTCCCTGCTCTTTAGAG 1611  
Db 361 aatgctatgaaatctcagaaatgagaaattcacagacgcttccctcgtctcttagag 420  
QY 1612 CCTGAGACCGGGGTAGTGTGATGTAAGTCTGAGATTTCTACGCGTAAAGTCCCAT 1671  
Db 421 cctgagaccggggtagtgtagctgtaactgctgagatctcccaacggatgagcccat 480  
QY 1672 TATGCTTTTCAAGCAATGTGAATTAAGTCTTGACAGAGACTGCAACAAAGTTGGAGAA 1731  
Db 481 tatgcttttcaagcaatgtgaaattagcttgacaggaactggaacaaagttggagaa 540  
QY 1732 TTTCTTGGGATTAATAAATTAATTAAGCTTGAACCTGAAAGAAAGAAATTTAGTTCT 1791  
Db 541 ttctctgggataaataaataataacctaagcctgaaagaagaataatctgctcc 600  
QY 1792 AATATCTTGTGGGGAATGCTGTAATGTGATCTTAATGACAGTTCAGAGAGACCAT 1851  
Db 601 aatatcttggtgggaatgctgtaaatgtcagatcctaataatgcaagttcaagagacat 660  
QY 1852 GCATCGAGCTTACAGACAGATCGAGGCTGTGATCGATGGAATTTGGAAATTTCTTCAT 1911  
Db 661 gcatcagacttacagacagatcgagggctgtgatacgtatggaatttcttccat 720  
QY 1912 GTATCTGCTCTCCGAAGCAATATAAGGTACGCTATAACACGGGTGGATTATGTTCTACT 1971





QY 852 AGAAGACATTCATTAATGGGCGGAGCCATTTATGCTAGGAACCTTAGACTCTTC 911  
 Db 961 caacatacatcgacaacagcgagagctactatgtaaaagctcgaaactgctc 1020  
 QY 912 AGGAGCTCCTACTATTTTATCAATATATATATATGCAAAATTCGCAAAATTTAGTG 971  
 Db 1021 cggagagctaacctcttctgtaagaatagtgatcaatggaggtacagctctctaaagtg 1080  
 QY 972 AGCTATGTCATTTGATAGTGGAGGAGATCAGTTTATGACAGAGAAAGAAATTTAC 1031  
 Db 1081 agccatagcatcgaaagatagtggaatgattatccgcgcatagtggtgacatgt 1140  
 QY 1032 ATTCCAGAGAAA-----CGGAGAGACTTACGTTTTTGATGGAGCAATCT 1079  
 Db 1141 cttttaaaggaaatacagctactctactactcctgggaagaatagaatgtagtgcct 1200  
 QY 1080 TTTTCAAAATGCTAAATTCCTGAAATTTACAGCGAGAAATGCTATTAATAATTTTA 1139  
 Db 1201 aggaagagtgcaagaatgacagcttggtctgctgtagagccatctactctta 1260  
 QY 1140 TGATCTATTACTTCTGAGCAGATGGGCTACCCAAATTAATATCAAGAGATCCTAA 1199  
 Db 1261 tgatccataactacagatgatacccaacagttacagatgctcttaaaatgtagac 1320  
 QY 1200 AATTAAG-----AGTACAGAGGACCATACCTTTTCTGGAGAAA----- 1240  
 Db 1321 tccggagatctcgcactacataatacagggaaacatcctcacaggaanaagtatc 1380  
 QY 1241 -AGACTTACAGAAAGATNCTTAGGATTTTAAATCTACAACTCCAGAACGCTCAACT 1298  
 Db 1381 aggaagagagcgagatctctaaatctactctgaagctactacagctgtaactc 1440  
 QY 1299 GTCGACGATTAATGATTTTAAAGAGGGGCCGAAGTCACAGTTTCAAAATTCACGCA 1358  
 Db 1441 ttaaggaagactctctcttaaaacatggagagctctcagactcagactaagctac 1500  
 QY 1359 GTCCTCAGATTCGATTTAGTATTAGATTAAGAACCAACTGATGCTTAGAGAGA 1418  
 Db 1501 acaagcgatctcgcctcgaatagcgttagaagaaactcctagaactgc-----ga 1554  
 QY 1419 CATTCGATCAGAGGCTCGCATAGATATATAGATAGCTTAACTCCTCAACAGACAG 1478  
 Db 1555 tactagcaccataaacaatttgctacttaacatcagctctatagaacgtgcaagaagc 1614  
 QY 1479 TGTATTAAAGCAAAACCCCAATTAACAGATATCCGTCAGGACTTATTAAGACTTAT 1538  
 Db 1615 aaaaatagaacccaagctacgtaaaaatctgactcttctcgaaacatctactat 1674  
 QY 1539 CTCGCTACTGGCAATGCCATGAGATCTCAGATGAGAAATTCACAGAGCTCCCTCT 1598  
 Db 1675 ggaacggaagcgacagttttagaaatcactagtttaagaatccctcagctcagacat 1734  
 QY 1599 GCTCTCTT---AGAGCTGAGCCGCGGGGTAGTGTGACTGATAGCTGAGATTTCT 1655  
 Db 1735 cttaaggtctaaagctcttgaaactgtaacaagcaacgagctgactccagatctacta 1794  
 QY 1656 ACCGTAAGTCCCAATTAAGTGTTCAGAGCAATTTG---AAATTAAGCTTGACAGAAC 1712  
 Db 1795 gggtagaatactcattacagctatcaggaacttggggccaattggttggggaagc 1854  
 QY 1713 TGGAAACAAGTTGGAATTTCTTGGGGAATAAATAATTAAGCTTAGACTGAGAAA 1772  
 Db 1855 ggccttaagcaactcctcaactcagactaaactgctatattccctcagcg 1914  
 QY 1773 AGAAGCAAAATTTAGTCTTAATATCTTGGGGAATGCTGTAATGTCAGATCTTAAT 1832  
 Db 1915 tatcgctctcttagtccctaatagtctatgaaatgcatlataatagctctctccca 1974  
 QY 1833 GCAAGTTCAAGAGACCCCATGAGCTTACAGACAGATGAGGGCTGTGATCGATGG 1892  
 Db 1975 ttatctatgagagctcaaacgaaggttgcagggagacgctcttcttgggtgctg 2034  
 QY 1893 AATTGGGAATTTCTTCATATCTGCTCCGAAAGCAATATTAAGTACCGTATTAACAG 1952

Db 2035 attatctactctctccatagaatagtlacaaaaacagcgaggttccatcttgag 2094  
 QY 1953 CGGTGATATGTTCTATCTGTAATAATGAGATCACACCTAACCACTATCTTGATGCG 2012  
 Db 2095 tggcggtatgcataggaagaaacctacactactgttccagataagttcttagtgctgc 2154  
 QY 2013 ATTTTCCCAACTCTTTTATGAGACAAAGACTATGCGGTTTCCAAACAGCAATACAGAT 2072  
 Db 2155 atttgcagctctcttggaaagataagactacttctgtagctaaagataaagttacagt 2214  
 QY 2073 GTATTTAGATTCGATCTATCATATATATACACCTCCCTAGGAAATTTTCCGTTATGC 2132  
 Db 2215 ctacggaagaaactctctattacacagcaacgaac-----tat 2254  
 QY 2133 TTCGCTAACCTTAATGTAACGTCGGGATTCCTCAAGAGGTTTCTTCAAAATCTCT 2192  
 Db 2255 atctctctcttgcgaactacagcccttgctgcttgcttctatgttccctacagatctc 2314  
 QY 2193 TATGATTTTCAATTTTGTGCTTATGCTCATGCCCAATATGATATGAAGACACTA 2252  
 Db 2315 gtctctcttccagaaaccttagc-taacccatcagatacagatctgaaaccaaigia 2373  
 QY 2253 CGCAAAATTTCCCTATGAGTGAAGAAACACTGAGAAACAAATTTGTTGGGCTATGAGTGG 2312  
 Db 2374 tacaacatacctactcgtttaaagaagcttggggaatgataagtttgccttagaatcgg 2433  
 QY 2313 AGGAGCATGCTCTTATGTTGATTTGAGAAAGGAAAGCTTTTCCAAAGTGCATCCCAT 2372  
 Db 2434 tggaaagaccccgatctgtcttagatgtaa---gtctctatcttgaagagatagccct 2490  
 QY 2373 TATGAACCTCAATTAATGATTTATGCTTATCAGGAGATTTTAAAGACAGCATGAGATGG 2432  
 Db 2491 catgaatctgcaatctgtctatcacatcagaaagtttaaaagaaaggaacgaagc 2550  
 QY 2433 CCGTAGATTTAGTAATGAGAGTTTAAACATGATTTTCTGATCTTACGATACGCTTTGA 2492  
 Db 2551 tctgtaatttggaagtagccgctcttgtaactctgctctacatccatccatccgacttga 2610  
 QY 2493 GAACTGGACATTTCTCAGAGATGACTCTATGACTTTAGTTTCTCTATATTCCTGATAT 2552  
 Db 2611 taagatcactgactgccaagaatgcaacgtaacatctactcttgatctatctgactc 2670  
 QY 2553 TTTCCGTAAGAGATCCCATGATGAGCTGCTGCTGATTAAGGAGAGACCTCGGCTGT 2612  
 Db 2671 tgtctgtagtaaccccgactgtagcaacacatcggaatagaggtgattcttggaaac 2730  
 QY 2613 TCCGCGACGACAGTATCAAGACATGCTTTGTAGGGAGTGAACGGGTCGGTATCACTT 2672  
 Db 2731 ctccgtagcaatttggcaagaacagctttagtccctcgtgcaggggaacatttgcct 2790  
 QY 2673 TAACGACTATACAGCTCTTATGTCGAGGAAGTATAGATGCGCCGCCCATGCTAGAA 2732  
 Db 2791 taactcaaatctttagaagccttagccaattctcttgaattgctggtggtcatctcgaa 2850  
 QY 2733 TTTAATATTAACGTGAGAGCAAAATTTCTTTTATGAAAGTTTCCATTG 2782  
 Db 2851 ttacaatgtagacttaggaagcaaataccatctctaagtcttagcttg 2900

## RESULT 6

AAX06820 standard; DNA; 2787 BP.  
 ID AAX06820 standard; DNA; 2787 BP.

AAX06820;  
 AC

26-APR-1999 (first entry)  
 DT

chlamydia pneumoniae surface exposed protein Omp8 DNA.  
 DE

Omp8; outer membrane protein 8; surface exposed protein; antigen;  
 KM

infection; diagnosis; vaccine; atherosclerosis; asthma; ss.  
 XX







Db 1217 TIKG-TESVTTSSQSG--DIGTISG-----TVEKATESLITQNSKIK 1259  
OY 555 AEDPSEYFQOGQWVNTTDT-----ATNTKEATA 585  
Db 1260 AT-----TGEANVTSATGTIGTISGNTVNVTA 1287

## RESULT 13

US-08-719-641-2  
Sequence 2, Application US/08719641  
Patent No. 6218141  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/719,641  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Belkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-625  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:  
US-08-719-641-2

Query Match 5.08; Score 238; DB 4; Length 1536;  
Best local similarity 21.1%; Pred. No. 2.6e-09;  
Matches 134; Conservative 99; Mismatches 235; Indels 168; Gaps 27;

OY 34 SFNGUID-----SFTFKTSATYTSLTGDF---FYEPKGTPLSDSCFQT----- 78  
Db 736 SFNGISVSGGSDVDFTLASSNVQRPVINSKYFNVSTGSSLR---FTSGSTKTGF 792  
OY 79 --TNNLTFGLNGSLTFEGITAGACAAASTANKNLTFSGFSLSDSSPSTVTITGQ 136  
Db 793 SIENDTLINATGNITLQVE-GTIDGMIGKGIIVAKKNTTFBGN-ITFGSKRAVIEIGN 850  
OY 137 GTLSAGGVNL-----ENIRLVVAGNESTADGAIKAGSFLITGTSDALFSNNSSFTK 191

Db 851 VTINNANVTLLIGSDPDFDHQKPL-----TIKKDYIINSGLTA 888  
OY 192 GGAITATTCARANTNTGTVRFSLNASTSGCAIDEGTSIIS-----NNKPLY 239  
Db 889 GGNIVINAGNLTVESNANFKATNTFENVGGLFEDKGSNLSIAKGAIRPDIDNSKNLS 948  
OY 240 FEGNAKTTTGAICNTKASGSPELIISNNKTLIFASNAETSGAIAHAK--LALSQGF 297  
Db 949 ITTNSSTYRTIISGNTINKNDLITNE-----GSDTEMQGVSGVSGEGTLTSSDKI 1003  
OY 298 TEELRNWVSATPKGA-----ISIDASELSIAETGNITFYRN----- 337  
Db 1004 -----NITKQITIKAGVGENSDSDATNNANLTITKELKLTQDINISGFNKAETIARDG 1058  
OY 338 ---TLTTGSDTTPKRNAINIGNSNGK---FTELR---AANKHTIFFDPITSECTSSDV 387  
Db 1059 SDLTIGNTNSAD-----GTNAKVTFNQVKDSKISADGHKVTLSKSVETSSNNT 1109  
OY 388 LKINNGSAGALNPYGTILFSGETLTADELKVAIDLKSSFTQPVLSGSKLLQKGTLE 447  
Db 1110 EDSDDNNAGL-----TIDAKNVYNNNTIHKAVSISATSGEITTKGTITIN 1156  
OY 448 STS-----FSQERGSLLG---MDSGT-TLSTAGSITITNLG-----INVDLSLKKQPVSL 494  
Db 1157 ATTGVEITATGTSILGIESSSGCVTLTATEGALAVNISGNTVTVANSGLTTLTLAGS 1216  
OY 495 TAKGASNKVIVYSGKLNLDIEGNIYESHMFSDQLFSLKITVDADVDPNIDISLIVP 554  
Db 1217 TIKG-TESVTTSSQSG--DIGTISG-----TVEKATESLITQNSKIK 1259  
OY 555 AEDPSEYFQOGQWVNTTDT-----ATNTKEATA 585  
Db 1260 AT-----TGEANVTSATGTIGTISGNTVNVTA 1287

## RESULT 14

US-08-614-377A-7  
Sequence 7, Application US/08614377A  
Patent No. 5976864  
GENERAL INFORMATION:  
APPLICANT: Smilt, John  
APPLICANT: Bingle, Wade H.  
APPLICANT: No. 5976864ellini, John F.  
TITLE OF INVENTION: EXPRESSION OF  
TITLE OF INVENTION: HETEROLOGOUS  
TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson PC  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/614,377A  
FILING DATE: 12-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/194,290  
FILING DATE: 09-FEB-1994  
CLASSIFICATION: 435  
APPLICATION NUMBER: US 07/895,367  
FILING DATE: 09-JUNE-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsao, Y. Rocky



```

Db 747 TGTFTVTLANATGTSVFNLLTSSAALAGVVALAGVE-----T 786
OY 537 VD-ADVDTN---VDISLIPVPAED--PNEYGOGQWNNMTDTATNTKEATATWTK 589
Db 787 VNIATDTNTTAHYDTLTQAATSAKSIYVTGNAGL---NLNTGNTAVTSFDSAV--- 839
OY 590 TGEVPSPERKSALVONTLMGVFTDIRSLQQLVEIGATGMEHKOGFPWSSMTNFLHKTGDE 649
Db 840 TGTAPAVTFVSA---NTTVGEVVTIR-----GGAGADSLTG----- 872
OY 650 NRKGRHTSGGVYIGS-----AHTPKDDLFT 676
Db 873 -----SATANDTIIGAGADTLVYTGTDFT 899

```

Search completed: October 2, 2001, 03:27:12  
 Job time: 7194 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 00:28:56 ; Search time 18674.8 Seconds  
(without alignments)  
2331.571 Million cell updates/sec

Title: US-09-446-677B-3  
Perfect score: 2815  
Sequence: 1 ATGAATGCAATTTCTG.....CTCATGTCACAAATTCG 2815

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_v1:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v11:\*  
59: gb\_v12:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_ro1:\*  
95: gb\_ro2:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2815	100.0	2815	9 A81829	A81829 Sequence 3
2	2815	100.0	17280	2 CPN133034	AJ133034 Chlamydia
3	2813.4	99.9	12676	1 AE002192	AE002192 Chlamydia
4	2813.4	99.9	300650	2 AP002546	AP002546 Chlamydia
5	2803	99.6	10757	1 AE001628	AE001628 Chlamydia
6	2742	97.4	3000	9 A81851	A81851 Sequence 25
7	2742	97.4	6030	2 CPOMP54	AJ001311 Chlamydia
8	670.6	23.8	2757	9 A81837	A81837 Sequence 11

QY	61	TCACGCTGTTTTGCTGCACACGCTGGAATAATATGAGCCCTCTGTTACTCTTTGACGGAAGT	120
Db	61	TCACGCTGTTTTGCTGCACACGCTGGAATAATATGAGCCCTCTGTTACTCTTTGACGGAAGT	120
QY	121	ACTAACACAGGCACCTTATCTCCTAAAAATATGCACTACTGGAATAGACTATCTTGACA	180
Db	121	ACTAACACAGGCACCTTATCTCCTAAAAATATGCACTACTGGAATAGACTATCTTGACA	180
QY	181	GGAGATTTAACTCTGCAGAAACCTTGGGGGATTCCGCACCTTTAACGAAGGGTGTCTTCT	240
Db	181	GGAGATTTAACTCTGCAGAAACCTTGGGGGATTCCGCACCTTTAACGAAGGGTGTCTTCT	240
QY	241	GACACGTACGGAACTCTTAAGCTTTGGCGGTAAAGGGGACACACTCTTTGTTTTAAATAT	300
Db	241	GACACGTACGGAACTCTTAAGCTTTGGCGGTAAAGGGGACACACTCTTTGTTTTAAATAT	300
QY	301	AAGCTATGTCGTGAAGGCGCAGCACTTCTGTATCAACTATATAAAATCTGTGCTATACA	360
Db	301	AAGCTATGTCGTGAAGGCGCAGCACTTCTGTATCAACTATATAAAATCTGTGCTATACA	360
QY	361	GGATTTTTCGAGTCTTACTTTCTTAAGCGGCCCATCATCGGTAATCACAMACCCCTCAGGA	420
Db	361	GGATTTTTCGAGTCTTACTTTCTTAAGCGGCCCATCATCGGTAATCACAMACCCCTCAGGA	420
QY	421	AAAGGTCAGATTAAATGTGGAGGGGATCTTTCATTTATATAACAATGGAACATATTATTTT	480
Db	421	AAAGGTCAGATTAAATGTGGAGGGGATCTTTCATTTATATAACAATGGAACATATTATTTT	480
QY	481	AAACAAGTTACTCTGAGGAAATAATGGCGGAGCCATTTCTACCAAGATCTTTCTTTGAAA	540
Db	481	AAACAAGTTACTCTGAGGAAATAATGGCGGAGCCATTTCTACCAAGATCTTTCTTTGAAA	540
QY	541	AACAGCACGGGATCGATTTCTTTTGAAGGAAATATCGAGCCGCAACAGGGAAAAAAGT	600
Db	541	AACAGCACGGGATCGATTTCTTTTGAAGGAAATATCGAGCCGCAACAGGGAAAAAAGT	600
QY	601	GGGGCTATTGTGCTACTGGTACTGTAGATTATTCAAATATATACGGCTCTACCCCTTTC	660
Db	601	GGGGCTATTGTGCTACTGGTACTGTAGATTATTCAAATATATACGGCTCTACCCCTTTC	660

QY	661	TCGAACATATTTGCTGAAGCTGCAGGGTGAAGACTTAAATATGACAGGAAACTGTACATT	720
	661	TCGACCATATTTGCTGAAGCTGCAGGGTGAAGACTTAAATATGACAGGAAACTGTACATT	720
Db	661	TCGACCATATTTGCTGAAGCTGCAGGGTGAAGACTTAAATATGACAGGAAACTGTACATT	720
QY	721	ACAGGGAATACGCTCTTGTATTTTCTGAAAATAGTGTGACAGCGACCGACGAAATGGA	780
	721	ACAGGGAATACGCTCTTGTATTTTCTGAAAATAGTGTGACAGCGACCGACGAAATGGA	780
Db	721	ACAGGGAATACGCTCTTGTATTTTCTGAAAATAGTGTGACAGCGACCGACGAAATGGA	780
QY	781	GGAGCTCTTCTGAGAGATGCCGATGTTACCATATCTGCGAATTCAGATGTACACTTCTCA	840
	781	GGAGCTCTTCTGAGAGATGCCGATGTTACCATATCTGCGAATTCAGATGTACACTTCTCA	840
Db	781	GGAGCTCTTCTGAGAGATGCCGATGTTACCATATCTGCGAATTCAGATGTACACTTCTCA	840
QY	841	GGAAACCAAGCTGACGTAAATGCGCGAGCCATTATGCTAAGAAGCTGTACACTGGCTTC	900
	841	GGAAACCAAGCTGACGTAAATGCGCGAGCCATTATGCTAAGAAGCTGTACACTGGCTTC	900
Db	841	GGAAACCAAGCTGACGTAAATGCGCGAGCCATTATGCTAAGAAGCTGTACACTGGCTTC	900
QY	901	GGGGGGGGGGGGGTATCTCTTTTCTAACATATYATGCCAAGTACCACTGCAGGTAA	960
	901	GGGGGGGGGGGGGTATCTCTTTTCTAACATATYATGCCAAGTACCACTGCAGGTAA	960
Db	901	GGGGGGGGGGGGGTATCTCTTTTCTAACATATYATGCCAAGTACCACTGCAGGTAA	960
QY	961	GGTGGACCACTTCTATACAGGCGACGTGGAGAGTGTCTTTTCAGGAGAAAGCGGGGAC	1020
	961	GGTGGACCACTTCTATACAGGCGACGTGGAGAGTGTCTTTTCAGGAGAAAGCGGGGAC	1020
Db	961	GGTGGACCACTTCTATACAGGCGACGTGGAGAGTGTCTTTTCAGGAGAAAGCGGGGAC	1020

Db 1021 ATTACCTCATGGGATGCCATTGTTCGCACTACACCAACAACCAAAAAGAAATCT 108

Db 1081 ATTGACATGAGGATCTACTGCAAAAGATCAGCAATTACGTGCATATCTGGCATAGCATC 114



```

|||||
Db 1141 TTTTTCATGATCCGATTAAGTCTGATATACGGCTGGGATTTCTACAGATTAATATCTC 1200
|||||
Qy 1201 AATAAGGCTGATGAGGATTAATAGTATATAGTGGGCTGATTTTCTTGTGTGAA 1260
|||||
Db 1201 AATAAGGCTGATGAGGATTAATAGTATATAGTGGGCTGATTTTCTTGTGTGAA 1260
|||||
Qy 1261 AAGCTCTGAGATGAAGCAAAAGTTGCACAACTTCTAGCTTAACAGCT 1320
|||||
Db 1261 AAGCTCTGAGATGAAGCAAAAGTTGCACAACTTCTAGCTTAACAGCT 1320
|||||
Qy 1321 GTAACCTTAATGAGGAAATTTAGTACTTAACGCTGCTCAGCTCAGTACGAAAGC 1380
|||||
Db 1321 GTAACCTTAATGAGGAAATTTAGTACTTAACGCTGCTCAGCTCAGTACGAAAGC 1380
|||||
Qy 1381 TTTACTGACAGCGGGGCTCTCTGTTATATGATGGCGGCACACGTTAAAGCAAGT 1440
|||||
Db 1381 TTTACTGACAGCGGGGCTCTCTGTTATATGATGGCGGCACACGTTAAAGCAAGT 1440
|||||
Qy 1441 ACAGAGGAGTCACTTAACAGGCTTTCCATCTCTGAGACCTTTAGGCGGAGGTAG 1500
|||||
Db 1441 ACAGAGGAGTCACTTAACAGGCTTTCCATCTCTGAGACCTTTAGGCGGAGGTAG 1500
|||||
Qy 1501 AAAGTTGTAATGCTGCTCTCTGAGCAAGTAAATGAGCCTTAGTGTCCGATTTCT 1560
|||||
Db 1501 AAAGTTGTAATGCTGCTCTCTGAGCAAGTAAATGAGCCTTAGTGTCCGATTTCT 1560
|||||
Qy 1561 CTTTGGATACCAAGGATATGTAATGAAATACAGCATTAAGGAAACTCAAGACTTT 1620
|||||
Db 1561 CTTTGGATACCAAGGATATGTAATGAAATACAGCATTAAGGAAACTCAAGACTTT 1620
|||||
Qy 1621 TCATTGTCAGGCTCTGCTCTGAGTACGCAACACTACAGATTTCCAGCGGTTCT 1680
|||||
Db 1621 TCATTGTCAGGCTCTGCTCTGAGTACGCAACACTACAGATTTCCAGCGGTTCT 1680
|||||
Qy 1681 ACAGTAGCACTCTCTAGGCACTATGATGATCAGTACTTGGGAAAGTCTGGGTTGAT 1740
|||||
Db 1681 ACAGTAGCACTCTCTAGGCACTATGATGATCAGTACTTGGGAAAGTCTGGGTTGAT 1740
|||||
Qy 1741 GATACCGGACACTCCCAAGACTTAAGACGACATTAAGCTTGGACCAATACAGGCTAC 1800
|||||
Db 1741 GATACCGGACACTCCCAAGACTTAAGACGACATTAAGCTTGGACCAATACAGGCTAC 1800
|||||
Qy 1801 CTTCGATCTCTGAGCGCAAGGACTTACTTCTAATACCTTTGGGATCTTTTCA 1860
|||||
Db 1801 CTTCGATCTCTGAGCGCAAGGACTTACTTCTAATACCTTTGGGATCTTTTCA 1860
|||||
Qy 1861 GACATCCAGGATTCAGAGTGTATAGAGAGAGTGTGACTTCTTTCAGATCGA 1920
|||||
Db 1861 GACATCCAGGATTCAGAGTGTATAGAGAGAGTGTGACTTCTTTCAGATCGA 1920
|||||
Qy 1921 GGCCTTGGGCTGGGGAGTGCCTAATTTCTTACATTAAGTAAAGGAGGAAAGCGC 1980
|||||
Db 1921 GGCCTTGGGCTGGGGAGTGCCTAATTTCTTACATTAAGTAAAGGAGGAAAGCGC 1980
|||||
Qy 1981 AAATACGCTCAATATCTGGTGAATGCTATGAGAGTGCAGGCAAACTTGTTCGAA 2040
|||||
Db 1981 AAATACGCTCAATATCTGGTGAATGCTATGAGAGTGCAGGCAAACTTGTTCGAA 2040
|||||
Qy 2041 AACTTAATGAGCTTGGCTTTTGGCAACTTTTGGTAGCGATTAAGATTTCTTATGCTCT 2100
|||||
Db 2041 AACTTAATGAGCTTGGCTTTTGGCAACTTTTGGTAGCGATTAAGATTTCTTATGCTCT 2100
|||||
Qy 2101 AAAAATATATGATATCTATGATGAGAGGCTTCTATATCCAAACATTTACAGATGAT 2160
|||||
Db 2101 AAAAATATATGATATCTATGATGAGAGGCTTCTATATCCAAACATTTACAGATGAT 2160
|||||
Qy 2161 GGGTTCATAGTGTCTCTATGATTAATCTTCTGGCTTGGAGTCAATTAACCCCTCGTT 2220
|||||
Db 2161 GGGTTCATAGTGTCTCTATGATTAATCTTCTGGCTTGGAGTCAATTAACCCCTCGTT 2220
|||||
Qy 2221 TTAGAAGGAGCTCGCTTATAGCCAGTCAATGATCTGAAGACAAATATATCTCG 2280
|||||

```

---

```

Db 2221 TTAGAAGGAGCTCGCTTATAGCCAGCTCAATGATCTGAAGACAAATATATCTCG 2280
Qy 2281 TATCTGAGGTGAAGGCTTCTTGGGGAAATATGCTTTTAACATGATCTTGGAGCTTCT 2340
Db 2281 TATCTGAGGTGAAGGCTTCTTGGGGAAATATGCTTTTAACATGATCTTGGAGCTTCT 2340
Qy 2341 TCTCATCTTATCCGAAATATCTGATGTTTGTATACCTTAAGCTTCAATCAAACTG 2400
Db 2341 TCTCATCTTATCCGAAATATCTGATGTTTGTATACCTTAAGCTTCAATCAAACTG 2400
Qy 2401 AATCGACCTATATACGTCAGACAGCTTCTCGAGAAAGGTACAGAAAGATCTTTT 2460
Db 2401 AATCGACCTATATACGTCAGACAGCTTCTCGAGAAAGGTACAGAAAGATCTTTT 2460
Qy 2461 GATGACAGCAACCTCTTCAATTAATCTTGGCTATAGGGGTGAAGTTGACAGTCTCT 2520
Db 2461 GATGACAGCAACCTCTTCAATTAATCTTGGCTATAGGGGTGAAGTTGACAGTCTCT 2520
Qy 2521 GATTGTATGACTTTTCTTATGATCTGACTTATCCCTATGTTCCGATCTATCCGCAAT 2580
Db 2521 GATTGTATGACTTTTCTTATGATCTGACTTATCCCTATGTTCCGATCTATCCGCAAT 2580
Qy 2581 GATCCCAATGCACTACAGCACTTGTATACAGCGAGCCTTGGGAAACTTATGCAAT 2640
Db 2581 GATCCCAATGCACTACAGCACTTGTATACAGCGAGCCTTGGGAAACTTATGCAAT 2640
Qy 2641 AACTTAGCAGACAGGCTTGCAGATGCTGAGGCACTACAGCTTCTCTCTATG 2700
Db 2641 AACTTAGCAGACAGGCTTGCAGATGCTGAGGCACTACAGCTTCTCTCTATG 2700
Qy 2701 TTTGAGTCTCGGCACTTGTCTTGAATGCTGATCCTCAAGGATTAATGTA 2760
Db 2701 TTTGAGTCTCGGCACTTGTCTTGAATGCTGATCCTCAAGGATTAATGTA 2760
Qy 2761 GATCTTGGGGTAAATTCCAAATCTAGAGGCTCTCATGCTCTCAGAAATCTG 2815
Db 2761 GATCTTGGGGTAAATTCCAAATCTAGAGGCTCTCATGCTCTCAGAAATCTG 2815

```

**RESULT 2**  
**CPN133034/c**  
**LOCUS**  
**DEFINITION**  
**ACCESSION**  
**VERSION**  
**KEYWORDS**

**SOURCE**  
**ORGANISM**  
**REFERENCE**  
**AUTHORS**  
**JOURNAL**  
**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**

**FEATURES**  
**source**  
**gene**  
**CDS**

1. 17280  
 /organism="Chlamydia pneumoniae"  
 /strain="VR1310"  
 /db\_xref="taxon:83558"  
 1. 841  
 /gene="omp12"  
 <1. 841

CPN133034 17280 bp DNA BCT 27-FEB-2001  
 Chlamydia pneumoniae GGAI-protein family cluster A, genes omp12  
 (partial), omp1, omp10, omp5, omp4, omp13 and omp14 (partial) and  
 ORF1, strain VR1310.  
 AJ133034.1 GI:4455881  
 hypothetical protein; omp10 gene; omp11 gene; omp12 gene; omp13  
 gene; omp14 gene; omp4 gene; omp5 gene; ORF1; outer membrane  
 protein; outer membrane protein 11; outer membrane protein 12;  
 outer membrane protein 14; outer membrane protein 4; outer membrane  
 protein 5.  
 Chlamydia pneumoniae.  
 Chlamydia pneumoniae.  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 1 (bases 1 to 17280)  
 Daugaard, L., Hjerno, K., Knudsen, K., Madsen, A.S., Christiansen, G.  
 and Birkebaard, S.  
 Unpublished  
 2 (bases 1 to 17280)  
 Boesen, T.  
 Direct Submission  
 Submitted (21-JAN-1999) Boesen, T., Department of Medical  
 Microbiology and Immunology, University of Aarhus, The Bartholin  
 Building, DK-8000 Aarhus, DENMARK  
 location/Qualifiers  
 1. 17280  
 /organism="Chlamydia pneumoniae"  
 /strain="VR1310"  
 /db\_xref="taxon:83558"  
 1. 841  
 /gene="omp12"  
 <1. 841

```

/gene="omp12"
/feature="putative"
/codon_start=2
/transl_table=11
/product="outer membrane protein 12"
/protein_id="CAB37067.1"
/db_xref="GI:4455882"
/translation="EDNIRYRHNSGCVAGITATTPAEDQLTFAPCOLAFDRNHTIG
KNKGDITYGSLFYHTEBGLFDINFLMGKATRAPWLSELSQIILPISDPKSEYLNHD
NHKITYTDSIILKGSWRNDACFADGLASLPEIVSYPLKEVEPVKVQYIYAHQD
FYRHAEGRAFNSELINVEIPGVTEERDSKEGTYDITLTIYDAYRRNKCQTS
LIASDANMAAYGTNLAROGFSYRAANHFQVNPHEIIFQGFAPFEVSRSSRNNTLSGK
FCF"
885..967
/gene="omp11"
/gene="omp11"
/gene="omp11"
885..3677
885..3677
/gene="omp11"
/feature="putative"
/codon_start=1
/transl_table=11
/product="outer membrane protein 11"
/protein_id="CAB37068.1"
/db_xref="GI:4455883"
/translation="MKIPLKLKLIISLTVPILISTATYCADASLSPDSDGSGST
FYRSTADANGTVLISGVNYINDAGKIALTCCEFTTGDLPFTGKGSFSEPNVD
AGSNAGAASSTADKALTFEGFENLSTIAFGTTVAAGKSTSSAGLANITDGTILF
SOVNSNEANNNGAIYTKLISGNTSITFTNSAKKLGAIYSMAASISGNTQL
VFNNGKGTGCGALGFEASSTIONSILFSGNATDAAGCAIYCEKTEGPTLTI
SGKSLIYFAENSSVTOGGACIAGHLAAGPILFSNNRCGTAAGCGAIAYADSS
LSISANGDITPLGNTLTSPASPTSRNALYLGSSAKITLRAAGQSIYFVPIASN
TTRASVLTINQDSDSPIDYSTIYSGEKLSDAKKADNFTSILKQPLAASGL
ALGNVELDNGFTQTEGSTLDPQGTKLADTEALSLKLVVDLSALBESKSVSIT
AGANKITLTLSPVFDSDSGNFESHITINQATLPQIVFTAAASDIYIDALLTSYV
OPEPHYOGHWEATWADSTAKSGMTWVTGYNPNPERRASVYVDSIMASEFTDR
TLQOIMTSQANSIYQOGLWASGTANFHKDKGTNOAFRHSXGYVGSASDEFNR
IFSVFCOLFGKDKLEFYENTSHNYALSILQHRAFLGLPMPSEGSITMDKIDPL
ILNAOISYTRKNDMDRTYSYPEAGOSYNNNGALIELGSLALYLPKMPROQGYR
FLRQAYIRKQOFKESGAERAPDDGDLVNCSTPVGIRLEKTSDEKKNPEISLAYI
GDVYRNPNRSRTSLWVSGASWTSLCNKLARAFLASGSHLTLSPHVELSGEAYELR
GSAHIYNDCGLRYSF"
8837..3907
/gene="omp10"
/gene="omp10"
3827..6613
/gene="omp10"
3827..6613
/gene="omp10"
/feature="putative"
/codon_start=1
/transl_table=11
/product="outer membrane protein"
/protein_id="CAB37069.1"
/db_xref="GI:4455884"
/translation="MKSLSHWPILISSSLALPLSNFSAFAVVEINLGPTNSPQCT
YTPAOTYNADGTYINLJGDVSTINAGSPALASCPKETGMLSGOGRHOYLOLNI
DAGANCTFTNTANKLSTFSGFSYLSLIQTNAITGAIKSTGACIQNSCYTRQ
NFSDNGALOGSSISLSLNPULTFAKNATORGALYSTGGITINNTLNASFSEMT
NANGGAIYEAESFISNKAISGFINSVATGATGAIYCSSTSAKPVLTALSDNE
LNFIGNAITSGAIYTDNLVLSGGPTLTKNNSAIDTAPLGGALIAAGSLSISA
LGDDITPEGNTYVKAASSQOTTRNSINIGTNAKIVOLRASQGNITYFVDPITTSIT
AALSDLANLNGPLAGNPAIOGIVSSEKLSAEAAENLKSITIOQPTTLAGGOLS
LKSGVTLNARSFSQSPGSTITLMDAGITLLETADGITTNNLVANDSLKTKKALKAQ
ASQVTLISGLSLVDPSSGVYEDVSMNPNQVFCILTLTADDPANIHITLDAAPLEKN
PIHMGYOGWALSMOEDTATSKRAALTWTKGTGNPNPERGTLVANTLGLASFEVDS
IOOLVATKVRQOETRGICWEGSINFEKHDSIKINGFRHISAGYVVGATTTLASDL
ITAPCOLGKRDHPIKNRKASAVASLSHLQHLATLSSPSILRYLPGSSRQDPLVD
AQSIYISKRMKTYTQAPKRGSSWYNGCCALIELASLIPHTALSHGELPHAYEPFK
VEASYIHQDSFERNTLVNPSDGLINISVITGITFEFRSRRERSATYIYIYAAD
VYRNNDCTTALLINNTSMKRTGTNLSRQIGRAGIEYAFSPNLVSTLSMEIRGS

```

```

CDS
8837..3907
/gene="omp5"
complement(8586..11372)
/transl_table=11
/product="outer membrane protein 5"
/protein_id="CAB37071.1"
/db_xref="GI:4455886"
/translation="MKSQFSWLIVLSSTLACFTCSNYFAATAENIGPSDFDGTWNG
TYPKNTGTGIDTYITGDTITLONLGPASALTRKCFSDTETSISFAGKGYSLFNILKS
SABGAASVYTDKNLSTIGFSSSLTFLPABRSYVITTPSGKAVYCGGDLTDDNGITLF
KOQYCENGGAIYSTKNLSLKNSTGSISFEBNKSATGKKGAIKATGVDYTNNTAP
LFENNNIAEAAGCAINSTGCTIYGNLSVSESNVATAGNGALSGADYITSGNS
VTFSGQAVANGCAIYAKKLTLASGGGGISFENNIVOGTLAGNGAISLAEBSL
SABAGDITTEGNAIVATTIPOTRKNSIDISTGKINLRAISGHSIFYDPIYANAA
SDTDITLNKAKADNGSTDGSIYVSEKSEDEKAVADNLITSLKQPTVLTAGNLV
LRGVTIDTGFQFOTAGSSVIMAGCTILAKSTEEVTLGASIPVDSIGBKXVYIAPSA
AKRNVALSGTILLDDNGARVENDLGRKQDSEFYVSLASGRTFTTDDVPVYIAPY
HGYQGTGWTGWDVDRATPKITYTATLAWNTGYLNPNEQGLVPSNLGVSFSDIOA
IOGVIRSAALTLSDGFMAAGVANFLDKDKERKRRKSGGVAIGGAQOCSNL
ISPAFOCLFSDSDPFLVAKNHTDYGAFYIQHITGCSGIGCLDLPGSMKHPDY
LEBQLAYSHVNDLKTRTYAPPEVKSQGNANMAGASHSYPEYLHCFDYYAYI
KMLITYIRODSFSEKGEGRSPDSNLFNLSLIGVKEFEISQCNDFSYDLTASYPD
LIRNDRKCTTAIYVSGASWETVANNLARAQLQVRAASHVAFSMFEVLQCFEYVGS
SRIVNDLGGKFOF"
complement(11313..11372)
/gene="omp5"
11535..14321
/gene="omp4"
11535..11588
/gene="omp4"
11535..14321
/gene="omp4"
/feature="putative"
/codon_start=1
/transl_table=11
/product="outer membrane protein 4"
/protein_id="CAB37072.1"
/db_xref="GI:4455887"
/translation="MKTSLPWLIVSSYLAFLSCHLQSLANELLSPDSFNGNIDSGTF
TPKTSATYVLTGDDVFEYERKGTPLSDSCFQKOTDMLPFLGNSHLSITFEYDIAGTHA
GAASSTANKNLTFSGSLSLSPSSSTYTTTQOGLTSSAGVYENIRKLYVAGNFS
TAGGAIKAGSFLTLTSGDALSSNNSSSTGGKALVTAARLANNGVYRFLSNLAS
TSGGALIDEOTSLISNNKFLYFEGMAKATYGAIICNTKASGPELLISNNKTLTFL
VATSGGAIHAKKLASLSSGFTFLNNVSSATPKGALSIDASGELSLSAEGNITF
VRNLTUTTGSTDTIPKRNAINIGSNGKFTELRAKNHTIPEYDPTSGSTSDVLYLN
GSAGLANPQGTTLFSGEITLADLEIKVADMLKSPQVYRSLSGKLLLOKGVLEST
FSGDAGSLDMDSGTTLSTTASGITTNLGLINWDSGLKQPVSLTAKGASNKYVSGK
LNLIDEGNTLYESHMSHDDLSFLKITYVADYDVINDVSSLLPVPRDEPNSEYCGQ
QMVNNTWTDATATKATATWTKTGEVPSERSALVCNTLGVFTDIRLQDLVEIG
ATGMEHQGFVWSMNTFLKTKDSENKGRHRHSGGVVIGGSATTPRDLFTFAFCL
FARKDKOEFILAHNNSRYGTLFFPKHSITLOPVYLRIGRAKFSESAIEKPPRIPLAL
DVQVFSHSDNMEHTYSLPESGSSNCSICAGGIGLDLPVLSNHPLEFKFTIEM
KVEMYVVSQNSPFESSDGRGFSIGHLNLSITVGAKTFQGDIGDSTYDLSGFFSD
VYRNNDCTTALVMSPDMSKIRGGINLSRQAFILRGSNNYTYNCELFGHTAMELKS

```

SRVNVDPVGTGRLEP"  
14579..14644  
/gene="omp13"  
14579..16123  
/gene="omp13"  
14579..16123  
/gene="omp13"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB57073.1"  
/db\_xref="GI:4455888"  
/translation="MTLRNPLFCSALFLPAAYVYVHSDNGYNGAINNLSLEPK  
ITCPYRGTSTYFLDVPRIISNVKHDQDAGVFINRSGLFPMKNCNTFENIMREGSG  
AATSNRYGDTTLTSLNFSYLAFTSAPILPQGGATISLGVMIENSEEVPFCGYSWS  
SGAATPYPLGSKASRPVNLGNRLVLRDVSQDYGGAISTHNLJTLTRGSCPE  
NNHAYHDVNSNGAIAIPAGSISISVKSQDLFTFKWTASODQNTIHNSJHLSGAOF  
KNLRVASESVYEYDPISSHESHKIDLVINAEKGKETEYGTISFSLCLDHEVCAL  
NLTSTLIDVTLVLAGTSLSDGYTLQHSKOEASSTLTPSPGTLICSGDARVONLH  
ILIEDDNEFVPRVIRAEKDAIVLSLEKTKAFAYVMSVDPQPKAEFTPLLEILGP  
SPDSLIGETTLERIVQYTTENDAVRGFMSLSMEERYPPSLDKDRITPTTKTVFLTRNP  
ELTSTP"

Query Match 100.0%; Score 2815; DB 2; Length 17280;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2815; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAATCGCAATTTCTCGTGTAGTGTCTCTTCGACATTCGATGTTTACTAGTTGT 60  
Db 11372 ATGAATCGCAATTTCTCGTGTAGTGTCTCTTCGACATTCGATGTTTACTAGTTGT 11313

Qy 61 TCCACTGTTTTGCGAACGCTGAAATATAGGCCCTCGATAGCTTTCAGCGAAGT 120  
Db 11312 TCCACTGTTTTGCGAACGCTGAAATATAGGCCCTCGATAGCTTTCAGCGAAGT 11253

Qy 121 ACTAACACAGGACCTTACTCTCTAAATAGACTACTGGAATAGACTATATCTGACA 180  
Db 11252 ACTAACACAGGACCTTACTCTCTAAATAGACTACTGGAATAGACTATATCTGACA 11193

Qy 181 GGAGATATATCTGTGCAAAACCTTGGGATTCGGACGCTTTAACGAGGGTGTCTTCT 240  
Db 11192 GGAGATATATCTGTGCAAAACCTTGGGATTCGGACGCTTTAACGAGGGTGTCTTCT 11133

Qy 241 GACACTAGCAATCTTAAAGCTTTCGCGGTAAAGGGTACTACTTCTTTTAAATTT 300  
Db 11132 GACACTAGCAATCTTAAAGCTTTCGCGGTAAAGGGTACTACTTCTTTTAAATTT 11073

Qy 301 AAGTCTAGTCTGAGAGGCGACACTTCTGTACAACTGATTAATAATCTGCTTAACA 360  
Db 11072 AAGTCTAGTCTGAGAGGCGACACTTCTGTACAACTGATTAATAATCTGCTTAACA 11013

Qy 361 GGATTTTCGACTTACTTCTTTCAGCGGCCCATCATCGGTAATCACACCCCTCAGGA 420  
Db 11012 GGATTTTCGACTTACTTCTTTCAGCGGCCCATCATCGGTAATCACACCCCTCAGGA 10953

Qy 421 AAAGTGACGTAATGTGGAGGGATCTTACATTTGATTAACATGGAACATTTTATTT 480  
Db 10952 AAAGTGACGTAATGTGGAGGGATCTTACATTTGATTAACATGGAACATTTTATTT 10893

Qy 481 AAACAGATTTACTGTAGAAATGGCGAGCATTCTTACAGAAATCTTCTTTGAAA 540  
Db 10892 AAACAGATTTACTGTAGAAATGGCGAGCATTCTTACAGAAATCTTCTTTGAAA 10833

Qy 541 AACGACACGGGATGATTTCTTTGAAAGGATTAATGAGCGCAACAGGAAAAAGT 600  
Db 10832 AACGACACGGGATGATTTCTTTGAAAGGATTAATGAGCGCAACAGGAAAAAGT 10773

Qy 601 GGGGCTATTGTGTACGGTACTGTAGATTTTACAAATATATAGGGTCCACCTCTTC 660  
Db 10772 GGGGCTATTGTGTACTGGTACTGTAGATTTTACAAATATATAGGGTCCACCTCTTC 10713

Qy 661 TCGAACAATATTTGCTGAAGCTGACAGGTGAGCTATTAATAGACAGAACTGTACATT 720  
Db 10712 TCGAACAATATTTGCTGAAGCTGACAGGTGAGCTATTAATAGACAGAACTGTACATT 10653

Db 10712 TCGAACAATATTTGCTGAAGCTGACAGGTGAGCTATTAATAGACAGAACTGTACATT 10653  
Qy 721 ACAGGGAATFAGTCTCTTGTATTTTCTGAAAATFAGTGTGACAGCGACCGCAGAAATGGA 780  
Db 10652 ACAGGGAATFAGTCTCTTGTATTTTCTGAAAATFAGTGTGACAGCGACCGCAGAAATGGA 10593

Qy 781 GGAGCTCTTTCGGAGATGCCGATGTACATATGTGGGAATCAGAGTGAATCTTCTCA 840  
Db 10592 GGAGCTCTTTCGGAGATGCCGATGTACATATGTGGGAATCAGAGTGAATCTTCTCA 10533

Qy 841 GGAAACCAAGCTGATAGCTAATGGCGAGCCATTATGCTAAGAACTTACACTGGCTTC 900  
Db 10532 GGAAACCAAGCTGATAGCTAATGGCGAGCCATTATGCTAAGAACTTACACTGGCTTC 10473

Qy 901 GGGGGGGGGGGGATATCTCTTTTCTAACAAATATAGTCCAAAGTACCACTGACAGTAAT 960  
Db 10472 GGGGGGGGGGGGATATCTCTTTTCTAACAAATATAGTCCAAAGTACCACTGACAGTAAT 10413

Qy 961 GGTGAGCCATTCTATATCTGACACTGAGAGTGTAGTCTTTCAGCAGAAACAGGGGAC 1020  
Db 10412 GGTGAGCCATTCTATATCTGACACTGAGAGTGTAGTCTTTCAGCAGAAACAGGGGAC 10353

Qy 1021 ATTACCTTCAATGGGAATGCCATTGTGCACACTACACCAACAAACTACAAAAGAAATCT 1080  
Db 10352 ATTACCTTCAATGGGAATGCCATTGTGCACACTACACCAACAAACTACAAAAGAAATCT 10293

Qy 1081 ATTGACATAGGATCTACGCAAAAGTACAGAAATTTACGTGCAATCTGCGGATAGCATC 1140  
Db 10292 ATTGACATAGGATCTACGCAAAAGTACAGAAATTTACGTGCAATCTGCGGATAGCATC 10233

Qy 1141 TTTTTCCTACAGATCCGATTAACGCTATACGGCTGGGATTCACAGATACTTAAATCTC 1200  
Db 10232 TTTTTCCTACAGATCCGATTAACGCTATACGGCTGGGATTCACAGATACTTAAATCTC 10173

Qy 1201 AATAAGCTGATGACGATTAATAGTACAGATTATAGTGGGTGATGTTTCTGCTGAA 1260  
Db 10172 AATAAGCTGATGACGATTAATAGTACAGATTATAGTGGGTGATGTTTCTGCTGAA 10113

Qy 1261 AAGCTCTGAAAGTGAAGCAAAAGTTCACACAACTCCTCTACGCTGAAGCAGCT 1320  
Db 10112 AAGCTCTGAAAGTGAAGCAAAAGTTCACACAACTCCTCTACGCTGAAGCAGCT 10053

Qy 1321 GTAACCTTAACGCGAAATTTACTTAAACGTGTGTCACCTCGATACGAAAGC 1380  
Db 10052 GTAACCTTAACGCGAAATTTACTTAAACGTGTGTCACCTCGATACGAAAGC 9993

Qy 1381 TTTACTAGACCGCGGTTCTCTGTATTAATGATGCGGGCCAAAGCTTAAAGCAAGT 1440  
Db 9992 TTTACTAGACCGCGGTTCTCTGTATTAATGATGCGGGCCAAAGCTTAAAGCAAGT 9933

Qy 1441 ACAGAGAGGTACCTTAAACAGGTCTTCCATCTCTGTAAGCTCTTAAAGGAGGTAA 1500  
Db 9932 ACAGAGAGGTACCTTAAACAGGTCTTCCATCTCTGTAAGCTCTTAAAGGAGGTAA 9873

Qy 1501 AAAGTTGTAATGCTGCTTCGACAGCAAGTAAATAATGTAAGCCCTTAGAGTCCGATTC 1560  
Db 9872 AAAGTTGTAATGCTGCTTCGACAGCAAGTAAATAATGTAAGCCCTTAGAGTCCGATTC 9813

Qy 1561 CTTTGTGTAACCAAGGGAATGCTTATGAAAATCAGCACTTAGAAAAACTCAAGACTTT 1620  
Db 9812 CTTTGTGTAACCAAGGGAATGCTTATGAAAATCAGCACTTAGAAAAACTCAAGACTTT 9753

Qy 1621 TCATTTTGACAGCTCTGCTGCTGGGTACTGCAACACTACAGATGTTCCAGCGGTTCCT 1680  
Db 9752 TCATTTTGACAGCTCTGCTGCTGGGTACTGCAACACTACAGATGTTCCAGCGGTTCCT 9693

Qy 1681 ACAGTAGCACTCTACGACATATGATGATCAAGGTACTTGGGGAATFAGTGGTGTAT 1740  
Db 9692 ACAGTAGCACTCTACGACATATGATGATCAAGGTACTTGGGGAATFAGTGGTGTAT 9633

Qy 1741 GATACCGCAACACTCCAAAGACTAAGACAGCAGCATTAAGTGTGGACCAATACAGGCTAC 1800  
Db 9632 GATACCGCAACACTCCAAAGACTAAGACAGCAGCATTAAGTGTGGACCAATACAGGCTAC 9573

OY	1801	CTTCCGAATCCTGAGCCTCAAGACCTTTAGTCTCCAAATAGCCTTTGGGACATCTTTTCA	1860
Db	9572	CTTCCGAATCCTGAGCCTCAAGACCTTTAGTCTCCAAATAGCCTTTGGGACATCTTTTCA	9513
OY	1861	GACATCCAGCGATTCAGAGTGTCATAGAGAGAAGTGCTTTGACTCTTTGTTCAGATCGA	1920
Db	9512	GACATCCAGCGATTCAGAGTGTCATAGAGAGAAGTGCTTTGACTCTTTGTTCAGATCGA	9453
OY	1921	GGCTTCCTGGGCTGGGGAGATCGCCAAATTTCTTGATAAAGATTAACAAAAGGGGAAAAAGC	1980
Db	9452	GGCTTCCTGGGCTGGGGAGATCGCCAAATTTCTTGATAAAGATTAACAAAAGGGGAAAAAGC	9393
OY	1991	AAATACCGTCAATTAATGTGTTGATATGCTATTCGAGAGGTGCAGGCCAACAATTGTTCTGAA	2040
Db	9392	AAATACCGTCAATTAATGTGATATGCTATTCGAGAGGTGCAGGCCAACAATTGTTCTGAA	9333
OY	2041	AACTTAATTAACCTTGGCTTTGTCGCAACCTTTGGAGAGATTTGTAATGATGCTT	2100
Db	9332	AACTTAATTAACCTTGGCTTTGTCGCAACCTTTGGAGAGATTTGTAATGATGCTT	9273
OY	2101	AAAAATCATACTGATACCTATGACAGAGACCTCTATATCAACACATTTACAGATATAGT	2160
Db	9272	AAAAATCATACTGATACCTATGACAGAGACCTCTATATCAACACATTTACAGATATAGT	9213
OY	2161	GGGTTCATAGGTTCTCCTTAGATTAACCTTCCGGGCTCTTGAGTCATTAACAAACCCCTGCT	2220
Db	9212	GGGTTCATAGGTTCTCCTTAGATTAACCTTCCGGGCTCTTGAGTCATTAACAAACCCCTGCT	9153
OY	2221	TTGAGAGGCGACCTCGCTTTATGACCAACGTGATTAATGATCTGAAAGACAAATATATCGG	2280
Db	9152	TTGAGAGAGCGACCTCGCTTTATGACCAACGTGATTAATGATCTGAAAGACAAATATATCGG	9093
OY	2281	TATCCTGAGGGAAGGTTCTTGGGGAAATATGCTTTAAATGATATGTGGGAGCTCT	2340
Db	9092	TATCCTGAGGGAAGGTTCTTGGGGAAATATGCTTTAAATGATATGTGGGAGCTCT	9033
OY	2341	TCTCATTTCTTATCCTGATACCTGCAATGTGTTTGATATACCTATGCTCATATACCAATCG	2400
Db	9032	TCTCATTTCTTATCCTGATACCTGCAATGTGTTTGATATACCTATGCTCATATACCAATCG	8973
OY	2401	AATCGACCTTATACGTCAGGACGACCTTCTCGGAAAGATGACAGAAAGAAATCTTT	2460
Db	8972	AATCGACCTTATACGTCAGGACGACCTTCTCGGAAAGATGACAGAAAGAAAGAAATCTTT	8913
OY	2461	GATGACAGCAACCTCTTCAATTTATCTTTCCTATATAGGGGTGAATTGTAGAAAGTCTCT	2520
Db	8912	GATGACAGCAACCTCTTCAATTTATCTTTCCTATATAGGGGTGAATTGTAGAAAGTCTCT	8853
OY	2521	GATTGTATATGACTTTTCTTATATGATATGACCTTATACCTATATGTCCTGATCTTATCCGCAAT	2580
Db	8852	GATTGTATATGACTTTTCTTATATGATATGACCTTATACCTATATGTCCTGATCTTATCCGCAAT	8793
OY	2581	GATCCCAATATGCACTACAGCACTGTATATACAGCGAGCCCTTTGGGAAACTTATGCCAAT	2640
Db	8792	GATCCCAATATGCACTACAGCACTGTATATACAGCGAGCCCTTTGGGAAACTTATGCCAAT	8733
OY	2641	AACCTTAGACAGCAAGGCTTGCAAGTGGCGAGAGGAGCACTACAGCCTTCTCTCATG	2700
Db	8732	AACCTTAGACAGCAAGGCTTGCAAGTGGCGAGAGGAGCACTACAGCCTTCTCTCATG	8673
OY	2701	TTTGAAGTGTCTGGGCGAGTTGTCTTTTGAAGTTCGTGATCTCTCAAGGATTTATATATGA	2760
Db	8672	TTTGAAGTGTCTGGGCGAGTTGTCTTTTGAAGTTCGTGATCTCTCAAGGATTTATATATGA	8613
OY	2761	GATCTTGGGGGTAAAGTTCCAAATCTTAGAGAGGCTCTCTCATGTCTAGAAATCTTG	2815
Db	8612	GATCTTGGGGGTAAAGTTCCAAATCTTAGAGAGGCTCTCTCATGTCTAGAAATCTTG	8558

RESULT	3
AE002192	
LOCUS	
AE002192	12676 bp DNA BCT 30-MAY-2000

DEFINITION	Chlamydia pneumoniae AR39, section 25 of 94 of the complete genome.
ACCESSION	AE002192 AE002161
VERSION	AE002192.2 GI:8163405
KEYWORDS	
SOURCE	
ORGANISM	Chlamydia pneumoniae AR39.
REFERENCE	Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.
AUTHORS	1 (bases 1 to 12676) Read,T.D., Brumham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Debony,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M. Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
TITLE	Nucleic Acids Res. 28 (6), 1397-1406 (2000)
JOURNAL	20150255
MEDLINE	10684935
PUBMED	2 (bases 1 to 12676) Read,T.D., Brumham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Debony,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M. Direct Submission Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA On Jun 1, 2000 this sequence version replaced gi:7189226.
COMMENT	Location/Qualifiers
FEATURES	1..12676 /organism="Chlamydia pneumoniae AR39" /strain="AR39" /db_xref="taxon:115711" /note="synonym: Chlamydia pneumoniae AR39" complement(90..185) /gene="CP0300" complement(90..185) /gene="CP0300" /note="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AAE38157.1" /db_xref="GI:7189227" /translation="MFTSLGRSAIFYFTGSLKRLGEFLEIRK" complement(275..1819) /gene="CP0301" complement(275..1819) /gene="CP0301" /note="similar to GP:4376735; identified by sequence similarity; putative" /codon_start=1 /transl_table=1 /product="polymorphic membrane protein G family" /protein_id="AAE38158.1" /db_xref="GI:7189228" /translation="MTLRNPLTCSALFLPAAQVYVTHESDINGAINKSLERKITYPEGTSTFLDQVRSINVKHQEDQGVIRSGNLFPMGRNCFEHNLTGEGGAALSNRYGDTTLTISNFSYALFTSAPLIPGOGAIVSIGSMENSEVEFTCNSSWSGAIVHYVYLIGSKASPSVNLGSGNRVLYLRDVAISOOGAISTHNLTITRGPSCEENNAHYTPVNLNGGAIAIAPGSGISISVKSGLDFKMGTAODGNTIHNSIHDSGOAFKNRAVSESVYFYDPISSHESHKIDVLYNAPBGRKTYRGCTISGSLCIDPHYCAELNLTSTLIQDVTLAGGTLSLDGVTQLQHSFKQDSSTLTNSPOTLLCSGDARVONHLILIEDDNPVAVRAIDKDALVLEKLKAFAYVSVDFPQKEAFTLIPLELLGPSPFSLISGETTLETTQVTTENDAVRGFWSISWEYEPSPIDKDRRIITPKTKVPLTWNPETSTP"
gene	complement(207..4926) /gene="CP0302" complement(207..4926) /gene="CP0302" /note="similar to GP:4376733; identified by sequence
CDS	

similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="polymorphic membrane protein G family"  
/protein\_id="AAFC38159.1"  
/db\_xref="gi:7189229"  
/translation="MIVLFQYIDANSSILKKKSITMKRSTIPWVYSSVIAFSCLOSIL  
APELLSDPDSFNGNIDSGTFTPTKRSATYSLTGDVFFYEKGKTPPLSDSCFKKTTDNL  
LTFLENGHSLGFTIDACTHAGAASTAAKNLTFPSGSLSPSSPSYTVTKGQTVL  
SSAGGVANLENIKRLVYAGNFSTADGALIKGASFLITGSGPALSNNSSRKGGAIAKTN  
TAGGAIAANTGQVFLSLINASTSGALIDDETSILSNKFLFYEGNAKLTGCAICMGT  
KASGEPIELISNNKLTIPASVNAETGALIAKKTALSSGFEFTGLNNVNSAAPKCGT  
AISIDASELSLSATGCTNITFYRNTLTGTSSTDTPKRAINIGSNKFTELRAAKNNTT  
IFETDPTSEGSVDVLLKINGSGALPVGDTLLFSEYTLTADELVADLKSGTFOV  
PVSLSGGKLLDQKGVLTLESTFSESOEAGSLTQMDSGTLSTAGSITITNLGINDVSDG  
LKQPSLSPKAGSNKSVIYSGKLTNIDLEGNIEHMEHQDLFSLKLTVDADVDVNTV  
DLSLILPAPAEADPNSEYFQFGOMVNTTDPATNTKATATWTKTGFVSPERSAIVY  
CTMLWGVTDLRSIQQLAIVETAGTGEHKGQVWVSMNTELTCTDENKQPRHNSGXY  
VYGSAAHPKPKDILTFPACFHLFARDKDFIAHNNSRYTGTLFHKHSHTLQPOYLLK  
GNAKSESAIEKFEPEIPLADVOVSEFSPHNSRMETHTSLPEBEGSWSMNCAIAGGLG  
LDLPVLSNPHPLEKTFIPQMKVEKVVYQSQSPSSSDGSGFISGLNLNLSIPVGAK  
FYQGIQSGSYTLDLGFSPVSVYRNPNPSTATLVSPSMKIRGGINLSRQAFILRGSN  
NVVYNSCELFQHYAMELRGSSRNNDVGRKLAF"

gene  
CDS  
5026..7812  
/gene="Cp0303"  
5026..7812  
/gene="Cp0303"  
/note="similar to GP:4376729; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="polymorphic membrane protein G family"  
/protein\_id="AAF38160.1"  
/db\_xref="gi:7189230"  
/translation="MKSQFSLVLSSTLACEFSGSYPAATENIGSDSPGDSNTNG  
TYTPKNTTGCDYLTITGDTLQNLGDSALUKKCPSPDTBESLSPAGCYSLSPNTILF  
SNEGALSVYTDKNLSTLTFESFLPLAPSSVITTPSGKAVKCGGDTLPDNNQTLIF  
KQDYLEENGGAISTNMLKNTGSISEKGSATGKGALICATGTVITNNATPT  
LFSNNIAAGALINSTGCTTNTSTLSEKNSATVATAGGALSGDADVTLSGANS  
VFESENQAVNGAIYAKKLTLAGSGGGGISENNIYQGTAGGAGAIISLAAGESSL  
SAEADIDPFGNCAIYATPOTTKRNSIDGSTAKTITNRAISGHSIFEPDIPITNAA  
DSTDPLNKKADAGNSTDYSGSIYFSGEKSEDEKAYADNLTSLKQPVTLTACNVL  
KRGVLLDPRKGFQTAGSSVIMDACITLKASTEEVTLIGLSIPVSLDEGRKVVIAAAS  
ASKNVALSGTILDLNQNAYENHDLGKTQDFSVOLASLTATTTVPVAPVTAAPT  
HGYGTGAGMTWVDVDTASTPKTKTATLAWMTNGVLPENBERGGPLVPSNLSFSDIOA  
IUGVIGERASLTICSDRGFWAGVANAFLTDKXKKGKRRYRHSGGVATIGAAQNCSEML  
ISFACQIFGSKDPELVAKNNTDYACAFYIQHTBESGFIQCLLDLPGSMHKLPLY  
LEGQIAYSHSDKTEKTKTAYAEVKSQGNNAFNMMLGASHSHSPEYLHCFDTPYAPYI  
KLNLYIKHDSSEKGTGRSPDSNLFNLSPIGVFEEKSDSSYDLTSLSYVD  
LIRNDPKCTTALVLSIGASWETVANNLARAQALQVABGSHYAFSPWFVILQGFVEVRSS  
SRIVNDLGGKRFQ"

gene  
CDS  
8329..9444  
/gene="Cp0305"  
8329..9444  
/gene="Cp0305"  
/note="conserved hypothetical protein; identified by  
glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF38162.1"

Query Match	99.9%	Score 2813.4	DB 1	Length 12676
Best Local Similarity	100.0%	Pred. No. 0		
Matches 2814	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY 1	ATGAATCGCAATTTTCCGTGGTACTGCGCTCTTGACATTTGGCATTTGGCATTTGTTTACTAGTTGT	60		
DB 5026	ATGAATCGCAATTTTCCGTGGTACTGCTCTTGACATTTGGCATTTGGTTTACTAGTTGT	5085		
QY 61	TCCACTGTTTTTGTCTGCAACTGCTGAAATAATATAGGCCCTCTGATAGCTTTGACGGAAGT	120		
DB 5086	TCCACTGTTTTTGTCTGCAACTGCTGAAATAATATAGGCCCTCTGATAGCTTTGACGGAAGT	5145		
QY 121	ACTAACACAGCACCCTATCTCTCTAATAAATAGACTACTGGAATAGACTATCTCTGACA	180		
DB 5146	ACTAACACAGCACCCTATCTCTCTAATAAATAGACTACTGGAATAGACTATCTCTGACA	5205		
QY 181	GGAGATTAATACCTGCAAAACCTTGGGATTCGGAGTTTAAGCAAGAGGTTGTTTTCT	240		
DB 5206	GGAGATTAATACCTGCAAAACCTTGGGATTCGGAGTTTAAGCAAGAGGTTGTTTTCT	5265		
QY 241	GACACTACGGAATCTTTAAGCTTTGCGGTAAAGGGTCTCTCACTTCTTTTAAATATT	300		
DB 5266	GACACTACGGAATCTTTAAGCTTTGCGGTAAAGGGTCTCTCACTTCTTTTAAATATT	5325		
QY 301	AAGTCTAGTCTGAAGGCGCAGACCTTCTGTGTACACTGATATAAATCTGTGCTAACA	360		
DB 5326	AAGTCTAGTCTGAAGGCGCAGACCTTCTGTGTACACTGATATAAATCTGTGCTAACA	5385		
QY 361	GGATTTTGAGCTTACTTTCTTACGGCCCCCATCATGGTAATCACAAACCCCTCAGGA	420		
DB 5386	GGATTTTGAGCTTACTTTCTTACGGCCCCCATCATGGTAATCACAAACCCCTCAGGA	5445		
QY 421	AAAGTCGAGTTAAATGTGAGAGGGATCTTACATTTGATAACAATGAACTATTTTATTT	480		

|||||  
Db 5446 AAAGTCAGTAAATGTGGAGGATCTTACATTGATAAACAATGAGAACTATTTATTT 5505  
QY 481 AAACAATACACTGTGAGAAAAATGGCGGAGCCATTTCTACCAAGAACTTTCTTTGAAA 540  
Db 5506 AAACAATACACTGTGAGAAAAATGGCGGAGCCATTTCTACCAAGAACTTTCTTTGAAA 5555  
QY 541 AACAGCAGGGGATCGATTTCTTTTGAAGGGAATTAATCGAGCGCAGACAGGAAAAAGT 600  
Db 5566 AACAGCAGGGGATCGATTTCTTTTGAAGGGAATTAATCGAGCGCAGACAGGAAAAAGT 5625  
QY 601 GGGGCTATTTGTGCTACTGTGATCTGTAGATTTACAAATAATACGGCTCTACCTCTTC 660  
Db 5626 GGGGCTATTTGTGCTACTGTGATCTGTAGATTTACAAATAATACGGCTCTACCTCTTC 5685  
QY 661 TCGAACAATATTTGCTGAGAGCGAGGAGCTTAATATACAGAGAACTGTACATT 720  
Db 5686 TCGAACAATATTTGCTGAGAGCGAGGAGCTTAATATACAGAGAACTGTACATT 5745  
QY 721 ACAGGAAATACGTCCTTGTATTTCTGAAAAATAGTGTGACAGCGACCGCAGAAATGA 780  
Db 5746 ACAGGAAATACGTCCTTGTATTTCTGAAAAATAGTGTGACAGCGACCGCAGAAATGA 5805  
QY 781 GGAGCTCTTCTGAGAGATGCCGATTTACCATATCTGGGAATCAGAGTGTACTTCTCA 840  
Db 5806 GGAGCTCTTCTGAGAGATGCCGATTTACCATATCTGGGAATCAGAGTGTACTTCTCA 5865  
QY 841 GGAACCAAGCTGTAGCTAATTTGGGAGCCATTATGCTMAAGAGCTTACCTGGCTTC 900  
Db 5866 GGAACCAAGCTGTAGCTAATTTGGGAGCCATTATGCTMAAGAGCTTACCTGGCTTC 5925  
QY 901 GGGGGGGGGGGGATCTCTCTTTCTACAAATATAGTCCAAAGGTACACTGACGATTAAT 960  
Db 5926 GGGGGGGGGGGGATCTCTCTTTCTACAAATATAGTCCAAAGGTACACTGACGATTAAT 5985  
QY 961 GGTGGAGCCATTTCTATACCTGAGCTGGAGAGTGTAGTCTTTACAGAGAGCGGGAC 1020  
Db 5986 GGTGGAGCCATTTCTATACCTGAGCTGGAGAGTGTAGTCTTTACAGAGAGCGGGAC 6045  
QY 1021 ATTACCTTCAATGGGAATGCCATTTGTGCAACTACACCAAACTACAAAAAGAAATCT 1080  
Db 6046 ATTACCTTCAATGGGAATGCCATTTGTGCAACTACACCAAACTACAAAAAGAAATCT 6105  
QY 1081 ATTGACATAGGATCTACTGCAAAAGATCAGCAATTTACGTGCAATATCTGGCATAGCATC 1140  
Db 6106 ATTGACATAGGATCTACTGCAAAAGATCAGCAATTTACGTGCAATATCTGGCATAGCATC 6165  
QY 1141 TTTTCTTACGATCCGATTTACTGCTAATACGGCTGCGGATTTCTACAGATTTTAAATCTC 1200  
Db 6166 TTTTCTTACGATCCGATTTACTGCTAATACGGCTGCGGATTTCTACAGATTTTAAATCTC 6225  
QY 1201 AATAAGGCTGTGAGGTAATAGTACAGATTAATAGTGGGTCGATTTGTTTTCGGTGA 1260  
Db 6226 AATAAGGCTGTGAGGTAATAGTACAGATTAATAGTGGGTCGATTTGTTTTCGGTGA 6285  
QY 1261 AAGCTCTCTGAAGTGAAGCAAAAGTTGCAGACAACCTCACTTCTAGCGCTAGACAGCT 1320  
Db 6286 AAGCTCTCTGAAGTGAAGCAAAAGTTGCAGACAACCTCACTTCTAGCGCTAGACAGCT 6345  
QY 1321 GTTAACTTAACTGAGAAATTTAGTACTTAACGTGGTGTCACTCTCGATACGAAAGGC 1380  
Db 6346 GTTAACTTAACTGAGAAATTTAGTACTTAACGTGGTGTCACTCTCGATACGAAAGGC 6405  
QY 1381 TTTTACACAGACCGGGGTCCTCTGTATTAATGATGCGGGCACAAGGTTAAAGCAAGT 1440  
Db 6406 TTTTACACAGACCGGGGTCCTCTGTATTAATGATGCGGGCACAAGGTTAAAGCAAGT 6465  
QY 1441 ACAGAGAGAGTCACTTTAAACAGGTCTTTCATTCCTGTAGACTTTTAAAGGAGGTAAG 1500  
Db 6466 ACAGAGAGAGTCACTTTAAACAGGTCTTTCATTCCTGTAGACTTTTAAAGGAGGTAAG 6525  
QY 1501 AAAGTTGTAATGTGCTTCTGACAGCAAGTAAATATAGCCCTTAGTGCTCGATTCCTT 1560  
|||||

Db 6526 AAAGTTGTAATGTGCTTCTGACAGCAAGTAAATATAGCCCTTAGTGCTCGATTCCTT 6585  
QY 1561 CTTTGTGATTAACCAAGGAATGCTTATGAATAATCAGCACTTAGGAAAAACCAAGACTTT 1620  
Db 6586 CTTTGTGATTAACCAAGGAATGCTTATGAATAATCAGCACTTAGGAAAAACCAAGACTTT 6645  
QY 1621 TCATTTGTGACGCTCTCTGCTGGGTACTGCAACAACATACAGATGTTCCAGCGGTTCT 1680  
Db 6646 TCATTTGTGACGCTCTCTGCTGGGTACTGCAACAACATACAGATGTTCCAGCGGTTCT 6705  
QY 1681 ACACTAGCAACTCTAGGCACTATGGGTATCAAGTACTTGGGAAATGACTTGGGTTGAT 1740  
Db 6706 ACACTAGCAACTCTAGGCACTATGGGTATCAAGTACTTGGGAAATGACTTGGGTTGAT 6765  
QY 1741 GATACCCGAACACTCCAAAGACTAAGACAGCAGCATTAGCTTGGACCAATACAGGCTAC 1800  
Db 6766 GATACCCGAACACTCCAAAGACTAAGACAGCAGCATTAGCTTGGACCAATACAGGCTAC 6825  
QY 1801 CTTCCGAATCCTGAGCGTCAAGACCTTTAGTTCCTAATAGCCTTTGGGATCTTTTCA 1860  
Db 6826 CTTCCGAATCCTGAGCGTCAAGACCTTTAGTTCCTAATAGCCTTTGGGATCTTTTCA 6885  
QY 1861 GACATCCAGGATTTCAAGTGTGATAGAGAGAGAGTGTGACTCTTGTTCAGATCGA 1920  
Db 6886 GACATCCAGGATTTCAAGTGTGATAGAGAGAGAGTGTGACTCTTGTTCAGATCGA 6945  
QY 1921 GGCCTTGGGGGCGGAGAGTCCCAATTTCTAGATTAAGATAAGGAGGAAAAAGC 1980  
Db 6946 GGCCTTGGGGGCGGAGAGTCCCAATTTCTAGATTAAGATAAGGAGGAAAAAGC 7005  
QY 1981 AAATACGTCATAATCTGTGTGATATGCTATCGAGGTGACGCGCAACTGTGTTGAA 2040  
Db 7006 AAATACGTCATAATCTGTGTGATATGCTATCGAGGTGACGCGCAACTGTGTTGAA 7065  
QY 2041 AACTTAATACCTTTGCTTTGCAACTCTTGTGAGCGATAAGAAATTTCTTGTGCT 2100  
Db 7066 AACTTAATACCTTTGCTTTGCAACTCTTGTGAGCGATAAGAAATTTCTTGTGCT 7125  
QY 2101 AAAAAATACATCTATACCTATACAGAGCCCTTCTATATCCAAACATTTACAGAAATGAT 2160  
Db 7126 AAAAAATACATCTATACCTATACAGAGCCCTTCTATATCCAAACATTTACAGAAATGAT 7185  
QY 2161 GGGTTCATAGTGTGTCTTATGATTAACCTTCTGGCTTGTGAGTCAATAAACCCCTGTT 2220  
Db 7186 GGGTTCATAGTGTGTCTTATGATTAACCTTCTGGCTTGTGAGTCAATAAACCCCTGTT 7245  
QY 2221 TTAGAAGGCGAGCTCGCTTATAGCCAGCTCAGTATGATCTGAAGACAAAGTATAGTGC 2280  
Db 7246 TTAGAAGGCGAGCTCGCTTATAGCCAGCTCAGTATGATCTGAAGACAAAGTATAGTGC 7305  
QY 2281 TATCCTGAGGTGAAGGTTCTTGGGGAAATTAATGCTTTTAAATGATGTTGGAGCTTCT 2340  
Db 7306 TATCCTGAGGTGAAGGTTCTTGGGGAAATTAATGCTTTTAAATGATGTTGGAGCTTCT 7365  
QY 2341 TCTCATTTCTATCTGTAATACCTGATTTGTTGATACCTATGCTCCATCATCAAACTG 7425  
Db 7366 TCTCATTTCTATCTGTAATACCTGATTTGTTGATACCTATGCTCCATCATCAAACTG 7485  
QY 2401 AATCTGACCTTATATACCTGAGACAGCTTCTCGGAGAAAGGTACAGAGGAAGTCTTTT 2460  
Db 7426 AATCTGACCTTATATACCTGAGACAGCTTCTCGGAGAAAGGTACAGAGGAAGTCTTTT 7485  
QY 2461 GATACAGCAACTCTTCAATTTATCTTGGCTATAGGGGTGAAGTTTGAAGATCTCT 2520  
Db 7486 GATACAGCAACTCTTCAATTTATCTTGGCTATAGGGGTGAAGTTTGAAGATTTCTCT 7545  
QY 2521 GATGTATATGACTTTTCTTATGATCTGACTTATCTATGTTCTGTATCTTATCCGAAT 2580  
Db 7546 GATGTATATGACTTTTCTTATGATCTGACTTATCTATGTTCTGTATCTTATCCGAAT 7605  
QY 2581 GATCCCAAAAGCACTTACAGACTTGTAAATCAGGAGAGCTTGTGGAACTTATGCCAAT 2640  
Db 7606 GATCCCAAAAGCACTTACAGACTTGTAAATCAGGAGAGCTTGTGGAACTTATGCCAAT 7665  
|||||



QY	2641	AACTTAGCAGCAGCAGCCCTTGCAAGCGCTGAGGAGCAGTCACTACGCCCTTCCTCTAG	2700
Db	7666	AACTTAGCAGCAGCAGCCCTTGCAAGCGCTGAGGAGCAGTCACTACGCCCTTCCTCTAG	7725
QY	2701	TTTGAAGTCTCGCCAGTTGTCTTTGAAGTTCGTGCATCCCTCAGATTTATATGTA	2760
Db	7726	TTTGAAGTCTCGCCAGTTGTCTTTGAAGTTCGTGCATCCCTCAGATTTATATGTA	7785
QY	2761	GATCTTGGGGGTAACTTCATTTCTAGAGCGCTGCTCATGTCTCAGAAATTCGT	2815
Db	7786	GATCTTGGGGGTAACTTCATTTCTAGAGCGCTGCTCATGTCTCAGAAATTCGT	7840
RESULT	4		
LOCUS	AP002546/c		
DEFINITION	Chlamydia pneumoniae J138 genomic DNA, complete sequence, section 2/4.		
ACCESSION	AP002546		
AB033802	AB033782	AB033783	AB033784
AB033803	AB033804	AB033805	AB033806
AB033807	AB033810	AB033811	AB033812
AB033813	AB033814	AB033815	AB033816
AB033817	AB033818	AB033819	AB033820
AB033821	AB033822	AB033823	AB033824
AB033825	AB033826	AB033827	AB033828
AB033829	AB033830	AB033831	AB033832
AB033833	AB033834	AB033835	AB033836
AB033837	AB033838	AB033839	AB033840
AB033841	AB033842	AB033843	AB033844
AB033845	AB033846	AB033847	AB033848
AB033849	AB033850	AB033851	AB033852
AB033853	AB033854	AB033855	AB033856
AB033857	AB033858	AB033859	AB033860
AB033861	AB033862	AB033863	AB033864
AB033865	AB033866	AB033867	AB033868
AB033869	AB033870	AB033871	AB033872
AB033873	AB033874	AB033875	AB033876
AB033877	AB033878	AB033879	AB033880
AB033881	AB033882	AB033883	AB033884
AB033885	AB033886	AB033887	AB033888
AB033891	AB033892	AB033893	AB033894
AB033895	AB033896	AB033897	AB033898
AB033901	AB033902	AB033903	AB033904
AB033905	AB033906	AB033907	AB033908
AB033911	AB033912	AB033913	AB033914
AB033915	AB033916	AB033917	AB033918
AB033919	AB033920	AB033921	AB033922
AB033923	AB033924	AB033925	AB033926
AB033927	AB033928	AB033929	AB033930
AB033931	AB033932	AB033933	AB033934
AB033935	AB033936	AB033937	AB033938
AB033939	AB033940	AB033941	AB033942
AB033943	AB033944	AB033945	AB033946
AB033947	AB033948	AB033949	AB033950
AB033951	AB033952	AB033953	AB033954
AB033955	AB033956	AB033957	AB033958
AB033959	AB033960	AB033961	AB033962
AB033963	AB033964	AB033965	AB033966
AB033967	AB033968	AB033969	AB033970
AB033971	AB033972	AB033973	AB033974
AB033975	AB033976	AB033977	AB033978
AB033979	AB033980	AB033981	AB033982
AB033983	AB033984	AB033985	AB033986
AB033987	AB033988	AB033989	AB033990
AB033991	AB033992	AB033993	AB033994
AB033995	AB033996	AB033997	AB033998
AB033999	AB033999	AB033999	AB033999

gene  
CDS  
/gene="CPj0268"  
957. .1343  
/gene="CPj0268"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAA98478.1"  
/db\_xref="gi:8978642"  
/translation="WSQSSSTSTWEMKSFVPNMKPPPLSPISSEDEFIAYPP  
FLPLTIDEMAOANPGTSTPNVENGIDDLNPLLGQPNQONNANNPETSGSNPISLPA  
PERLETENSGEEOGSGQNNEDILIG"  
complement(1495. .2472)  
/gene="CPj0269"  
complement(1495. .2472)  
/gene="CPj0269"  
/codon\_start=1  
/transl\_table=11  
/product="dipeptidase"  
/protein\_id="BAA98479.1"  
/db\_xref="gi:8978643"  
/translation="WTIDMCDLISHPHCRKDAVRCSPQLISGVQVQCAIYVP  
HSRSPNDCKONSLEFFSPNQYPIDGLSYEEENGSSQKKSLSLRISLMSALD  
DTAPGATGLAKILHTKQGPLAYLGIYWKGNFRFGSEAPKRLSNGKYLIDMYLE  
GVPIIDSCSDKLADIIDIDYADKLPLNAYIASHNSNFRSYDHRNRIVDAHAKIYVR  
KGVGINIVRSYVGDSIGDLEKHVLAHNGIISYIYGSGPFYANDENPFNECCS  
AAHAYNLQDILHRIKSKKASIISSRAEKLKQYIYQVNPKITDKL"  
complement(2487. .3347)  
/gene="yw1c"  
complement(2487. .3347)  
/gene="yw1c"  
/codon\_start=1  
/transl\_table=11  
/product="SuA5 superfamily-related protein"  
/protein\_id="BAA98480.1"  
/db\_xref="gi:8978644"  
/translation="MPDCKAOTISLPEVMSAIHQGKIVALPDPYVGVFLSYASPA  
BERLYALMDRPSKRFALYVNSIEDIENISYPLSPAKKLAQFPFPAITILYKHNRP  
RPFKSTLFRIVDSVREIYDHGCTGLIGTSANISEPSSALTAQETIRADPHDCLTE  
DQPCSHGLESTVYASDPLIYIREGLISNYSYIAGTEAKIFHTSHAFKSHKIYV  
KNQEDLVFELFSGSLDFKGVCEHPKPKFYRLREALKKTPPSIVETIINTSPYEL  
PFELSPYIE"  
3656. 4387  
/gene="CPj0271"  
3656. 4387  
/gene="CPj0271"  
/codon\_start=1  
/transl\_table=11  
/product="lysophospholipase esterase"  
/protein\_id="BAA98481.1"  
/db\_xref="gi:8978645"  
/translation="MTDYSFFRRKIGNIEAIECPNODPIIILCHGYSGLADNLFF  
PSICGSEFLRPWIEPNGILPLENDPFSRACPLNLTALHESRLYAVNGVNOKEY  
DELPEQDLEPPEKALELLEILNLNPNVMEIISGSGCAIILATHTLSQNPYAGALF  
AGARLFNQGMEGLKQCAQVFPLOSHGEDEITLPHLGAHLNDLLLTQNGVYFRHG  
GHEIPSVFQMQVYVPMWIDPARG"  
complement(4359. .5255)  
/gene="dnaX\_2"  
complement(4359. .5255)  
/gene="dnaX\_2"  
/codon\_start=1  
/transl\_table=11  
/product="DNA polymerase III gamma and tau"  
/protein\_id="BAA98482.1"  
/db\_xref="gi:8978646"  
/translation="MHLEENQGEALILKRYVHOEVPAILLHGFTLPVLYQDAEOLA  
SEILLSSPGESEKVSOKIHPIYQFPPEKGRHSHIDLRGIRKQYIYSPFPAANKI  
YIIEADPMTLTAIASFLKVEEPPKHAIVILITAKVORLPKTIISRSLSIFEREK  
ILCEKRETSYFRAACEIYVTEEVSQIIEKSESTDKOVLDRKQVREMYVLELYRDY  
TLNGLKASALNYPEHVKETIOLPLPLDKVYLIVESACRSLNSSSAASVLEMAIO  
LVSIQYKEKELVSPGODLSN"  
complement(5246. .5866)  
/gene="tdk"  
complement(5246. .5866)





QY 1081 ATTGACATAGGATCTACTGCAAGATCAGAAATTAGCTGCATATCTGGGCATAGCATC 1140  
 |||||||  
 Db 206750 ATTGACATAGGATCTACTGCAAGATCAGAAATTAGCTGCATATCTGGGCATAGCATC 206691  
 QY 1141 TTTTCTACAGATCGGATTAAGTCTAATAGCGCTCGGATTTCTACAGATTAATTAATCTC 1200  
 |||||||  
 Db 206690 TTTTCTACAGATCGGATTAAGTCTAATAGCGCTCGGATTTCTACAGATTAATTAATCTC 206631  
 QY 1201 AATTAAGCTGATGAGGTAATAGTACAGATTAATAGTGGGTGATTTCTTTCTGGTGA 1260  
 |||||||  
 Db 206630 AATTAAGCTGATGAGGTAATAGTACAGATTAATAGTGGGTGATTTCTTTCTGGTGA 206571  
 QY 1261 AAGCTCTGAGATGAGCAAAAGTGCAGACAACTCACTTTAGCGTGAAGCAAGCT 1320  
 |||||||  
 Db 206570 AAGCTCTGAGATGAGCAAAAGTGCAGACAACTCACTTTAGCGTGAAGCAAGCT 206511  
 QY 1321 GTAACCTAATGAGGAAATTTAGTACTTAACGTGGTGTCACTCTCGATACGAAAGC 1380  
 |||||||  
 Db 206510 GTAACCTAATGAGGAAATTTAGTACTTAACGTGGTGTCACTCTCGATACGAAAGC 206451  
 QY 1381 TTTACTAGACCGCGGTTCTCTGTTATTAATGATCGCGGCACACGTTAAAGCAAGT 1440  
 |||||||  
 Db 206450 TTTACTAGACCGCGGTTCTCTGTTATTAATGATCGCGGCACACGTTAAAGCAAGT 206391  
 QY 1441 ACAGAGGAGTCACTTAACAGGCTTTTCCATTCCTGTAGACCTTTAGGCGAGGTAAG 1500  
 |||||||  
 Db 206390 ACAGAGGAGTCACTTAACAGGCTTTTCCATTCCTGTAGACCTTTAGGCGAGGTAAG 206331  
 QY 1501 AAGTTGTAATGCTGCTTCTGACAGCAAGTAAATAGTGGCTTATGCTCCATCTT 1560  
 |||||||  
 Db 206330 AAGTTGTAATGCTGCTTCTGACAGCAAGTAAATAGTGGCTTATGCTCCATCTT 206271  
 QY 1561 CTTTGGATTAACCAAGGAATGCTTATGAATATACGACTTGAAGAAATCTCAAGACTTT 1620  
 |||||||  
 Db 206270 CTTTGGATTAACCAAGGAATGCTTATGAATATACGACTTGAAGAAATCTCAAGACTTT 206211  
 QY 1621 TCATTGTGAGCTCTCTGCTGGTACTGCAACACTACAGATGTTCCAGCGGTTCT 1680  
 |||||||  
 Db 206210 TCATTGTGAGCTCTCTGCTGGTACTGCAACACTACAGATGTTCCAGCGGTTCT 206151  
 QY 1681 ACAGTACCAACTCTACGACTATGGGTATTCAGAGTCTGGGGAATGACTTGGGTTGAT 1740  
 |||||||  
 Db 206150 ACAGTACCAACTCTACGACTATGGGTATTCAGAGTCTGGGGAATGACTTGGGTTGAT 206091  
 QY 1741 GATACCGCAAGCACTCAAGAGCTAAGACGCACTTACTTTGACCAATACAGGCTAC 1800  
 |||||||  
 Db 206090 GATACCGCAAGCACTCAAGAGCTAAGACGCACTTACTTTGACCAATACAGGCTAC 206031  
 QY 1801 CTTCCGATCTCTGAGCTGTCAGGACTTTAGTCCTAATAGCCTTTGGGATCTTTTCA 1860  
 |||||||  
 Db 206030 CTTCCGATCTCTGAGCTGTCAGGACTTTAGTCCTAATAGCCTTTGGGATCTTTTCA 205971  
 QY 1861 GACATCCAGCGATTCAGGTGTATAGAGAGAGTCTTGACTCTTGTTCAGATCGA 1920  
 |||||||  
 Db 205970 GACATCCAGCGATTCAGGTGTATAGAGAGAGTCTTGACTCTTGTTCAGATCGA 205911  
 QY 1921 GGTCTTGGGCTCGGGAGTCCCAATTTCTTATGATTAAGAAAGGGAAGGCAAGC 1980  
 |||||||  
 Db 205910 GGTCTTGGGCTCGGGAGTCCCAATTTCTTATGATTAAGAAAGGGAAGGCAAGC 205851  
 QY 1981 AAATACCGTATTAATCTGGTGTATGATATGATGAGAGTCCAGGCAAACTGTCTGAA 2040  
 |||||||  
 Db 205850 AAATACCGTATTAATCTGGTGTATGATATGATGAGAGTCCAGGCAAACTGTCTGAA 205791  
 QY 2041 AACTTAATAGCTTGGCTTTGGCAACTCTTTGATGAGATTAAGAAAGATTTCTTACTGCT 2100  
 |||||||  
 Db 205790 AACTTAATAGCTTGGCTTTGGCAACTCTTTGATGAGATTAAGAAAGATTTCTTACTGCT 205731  
 QY 2101 AAAAATATATCTGATTAATGAGAGGCTCTTATATCAACACATTAAGATATGAT 2160  
 |||||||  
 Db 205730 AAAAATATATCTGATTAATGAGAGGCTCTTATATCAACACATTAAGATATGAT 205671

QY 2161 GGGTTCAATAGGTGTCCTTAGATAACTTCCTGGCTCTTGAGTCAATTAACCCCTGCT 2220  
 |||||||  
 Db 205670 GGGTTCAATAGGTGTCCTTAGATAACTTCCTGGCTCTTGAGTCAATTAACCCCTGCT 205611  
 QY 2221 TTGAAGGGCAGCTCGCTTATAGCACGTCAGTAAATGATCTGAAGACAAATATCTCG 2280  
 |||||||  
 Db 205610 TTGAAGGGCAGCTCGCTTATAGCACGTCAGTAAATGATCTGAAGACAAATATCTCG 205551  
 QY 2281 TATCTGAGCTGAAGGTTCTTGGGGGAATATGCTTTTAACATGATTTGGAGCTTCT 2340  
 |||||||  
 Db 205550 TATCTGAGCTGAAGGTTCTTGGGGGAATATGCTTTTAACATGATTTGGAGCTTCT 205491  
 QY 2341 TCTCATCTTATCTGATATACCTGATTTGTATACCTTATGCTCCATATCAATCAATG 2400  
 |||||||  
 Db 205490 TCTCATCTTATCTGATATACCTGATTTGTATACCTTATGCTCCATATCAATCAATG 205431  
 QY 2401 AATCTGACATATATACGTCAGACAGCTTTCGAGAGAAAGGTACAGAGAAATCTTTT 2460  
 |||||||  
 Db 205430 AATCTGACATATATACGTCAGACAGCTTTCGAGAGAAAGGTACAGAGAAATCTTTT 205371  
 QY 2461 GATGACAGCAACCTCTTCAATTTATCTTGGCTATAGGGGTGAAGTTGAGAGTTCTCT 2520  
 |||||||  
 Db 205370 GATGACAGCAACCTCTTCAATTTATCTTGGCTATAGGGGTGAAGTTGAGAGTTCTCT 205311  
 QY 2521 GATGTAATGACTTCTCTTATGATCTGACTTATCCTATGTCCTGATCTATCGCAAT 2580  
 |||||||  
 Db 205310 GATGTAATGACTTCTCTTATGATCTGACTTATCCTATGTCCTGATCTATCGCAAT 205251  
 QY 2581 GATCCCAATGCACTACAGCACTTGTATACGAGGAGCTTTGGGAACTTAAGCCAAAT 2640  
 |||||||  
 Db 205250 GATCCCAATGCACTACAGCACTTGTATACGAGGAGCTTTGGGAACTTAAGCCAAAT 205191  
 QY 2641 AACTTACAGACAGCGCTTTCAGAGTGTGTCAGAGCACTACAGCTTCTCTCATG 2700  
 |||||||  
 Db 205190 AACTTACAGACAGCGCTTTCAGAGTGTGTCAGAGCACTACAGCTTCTCTCATG 205131  
 QY 2701 TTTGAATGCTCGGCAAGTTTGTCTTGAAGTGTGATGATCTCAAGGATTTATATGTA 2760  
 |||||||  
 Db 205130 TTTGAATGCTCGGCAAGTTTGTCTTGAAGTGTGATGATCTCAAGGATTTATATGTA 205071  
 QY 2761 GATCTTGGGGTAATTCATTTCTAGAGCGTCTCATGCTCAGAAATCTG 2815  
 |||||||  
 Db 205070 GATCTTGGGGTAATTCATTTCTAGAGCGTCTCATGCTCAGAAATCTG 205016  
 RESULT 5  
 AE001628/c DNA BCT 01-DEC-2000  
 LOCUS AE001628 10757 bp  
 DEFINITION Chlamydia pneumoniae section 44 of 103 of the complete genome.  
 ACCESSION AE001628  
 VERSION AE001628.1 GI:4376730  
 KEYWORDS  
 SOURCE Chlamydia pneumoniae CwL029.  
 ORGANISM Chlamydia pneumoniae CwL029  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 REFERENCE  
 1 (bases 1 to 10757)  
 AUTHORS Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,  
 Olinger, J., Grimwood, J., Davis, R.W. and Stephens, R.S.  
 TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
 JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
 MEDLINE 99206606  
 PUBMED 10192388  
 REFERENCE  
 2 (bases 1 to 10757)  
 AUTHORS Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,  
 Grimwood, J., Davis, R.W. and Stephens, R.S.  
 TITLE Direct Submissions  
 JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University  
 of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
 FEATURES  
 source  
 1. /organism="Chlamydia pneumoniae CwL029"  
 /strain="CwL029"  
 /db\_xref="taxon:115713"

```
gene
106..2892
/ gene="pmp_9"
/ note="Cp00447"
CDS
106..2892
/ gene="pmp_9"
/ codon_start=1
/ transl_table=11
/ product="Polymorphic Outer Membrane Protein G/I Family"
/ protein_id="AADI8591.1"
/ db_xref="GI:4376731"
/ translation="MKSSILHPLISSSLALPLINSEFAAYVEINLGPINSFSGPCT
YTPRACTTNADGRTYINLGDVSTINNASPFLATPASCRTETGNSLSPGHGQYPLONT
DAGANCTFTYPAANKLSESGESTLSLQTNMTGATGALSTGACISQSNISCFEQ
NENGDGALQSSISLSLNPNTFANKNATOGKALYSTGTTINNTLMSASENT
AANNNGALYEASFSISNKAISFINSVATSGATGAIYCSSTAPKPVLLTASDNE
LNFIGNATISGATIDNLYVLSGGGTLEKNNASIDTAPLGLAIALSGSLSLA
LGGDIPREGNTVYKGAASSQTTNRSININNTAKIYOLASQGNITLYFDDPTTST
AALSDALNLPDLACNPAOCTIVSGEKLSEEAADNLKSTIOQPLTLAAGQIS
LKSQVTLVAKSFOSPGSTLLMDGTTLEADGTTINNLVANDSLKETKATLKAQ
ASQVTLISGLSLVDPSGNYEDVSNMNPVFSCLTLTADDPANILHTDLAADPLEKN
PIHMVYOGNMLSMOEDTATSKAATLTMTKTGNPNERGTLVANTLNGSEVDEN
IIOOVLATKVRQSOETRGICEGISNFKDSTKNGKFRHSAGYVYVATTLASDL
ITAPFOLPEKDRDHEFNKRAASAYASLHOLHATLSSPILRYLPSSSDQVLEED
AOSIYKSKMPTKTYTOQAPKGSNWDNDCALFELASLPHTALSHGLTHATPPRK
VASYIHQDSFKERNITLVRSFDSGLINVSPIGTFEFSRNERASTATYIYVAD
VYRNKPDCTTALLINMTSMKTTGTLNLSRQAGISRAIGFYAFSPULEVTNLSMEIRGS
SRSYNDLGRKFPF"
gene
complement (333..4348)
/ gene="yx1G_2"
/ note="Cp00448"
CDS
complement (333..4348)
/ gene="yx1G_2"
/ codon_start=1
/ transl_table=11
/ product="Hypothetical Protein"
/ protein_id="AADI8592.1"
/ db_xref="GI:4376732"
/ translation="MEMSPFOQPEOCHPVPVGSFLPESLTPARSDFEGRGIVYEQM
RVEDDAIRNLIRKQTEAGLIFPTDGFRRYSWDFPMWFHVDNRDRDNDDEIGY
LKDKISYKPFIEHEFEVTEFKRNKAQOTIPSPQSFHEMIFAPNLNTRKLYET
NOELIDDIYFYRYQIOLDLAAGCRNADQCATLIDIRASWYGVDSHDLRQELT
EOLFNLHNLVAKRDEDLFVSLHVCGRDYGAEFERSISTEIPFAKIDVDSYHYH
WALDDKYSGAAPLAYSGEKHYVCLGISNHCIDRQAVRSIRIYEAASYIPLERLS
LSQCGFASCEGHRTHEEOKKIAFVKETAEIKW"
gene
complement (485..6769)
/ gene="pmp_10.1"
/ note="Cp00449"
misc_feature
complement (485..7650)
/ note="Polymorphic Outer Membrane Protein G Family;"
contains frame shift"
complement (6715..7650)
/ gene="pmp_10.2"
/ note="Cp00450"
7813..10599
/ gene="pmp_11"
/ note="Cp00451"
7813..10599
/ gene="pmp_11"
/ codon_start=1
/ transl_table=11
/ product="Polymorphic Outer Membrane Protein G Family"
/ protein_id="AADI8593.1"
/ db_xref="GI:4376733"
/ translation="MKTSPWLVYSSVLAFSCHLQSLANBELSLPDSFNGNIDSGTF
TPKTSATYSLNDGVFEYBKGPPLSDSCFKQTNLQIFLGHSLHLETFGIDAGHA
GAASTTANKNLTFSGSLSPDSPTTYTTOGTLSSAGVNLNIRLYLVYASNS
TADGAIKASPELLTGTSGDALPSNNSSTKGAIAITTGAGARIANNVGYRELINAS
TSGAIDDBGCTSLSNKKFLYEGNAKATGGATCINTKASGSELLITSNKTLIFASN
VATSGGAIHAKKALSLSGFTEFLNNVSSATPKGASIDSGLSLSAEGNITFE
VNTLTITGSTDTPKRNAINIGNGKFTLPAKNTAIFYPDIITSEGSISDLKLTNN
GSAALNPYOGTILFSGETLADLKVADLUKSFPTOPSLSGKLLDKGVYLESTS
FSQEGALNPGSGTLLSTAGSITITTINGINDSGLKOPVSLTAKGASNKYIVSGK
LNIIDIEGNIYESHMFSDQLSLKITVDADVTVNDVDSISLIPVAEDPNSRYGQG
```

```
OMNVNNTTDTATNTEKATNTWTGVPSEPKRSALVCNTLKGVTDIRSLQOLVEIG
ATGWEHRQGFWSMTNETHLKHKTDENNKGRHISGGLVIGSAAHPEDDLLEPAFCHL
FARDKDCFIAHNNSRYTGTLFFKHSHTLOPQWYLRGRAKFSESALKEPREIPLAL
DVOVSFSDNRMEHTTSLPESGSSWNSIGIGLIDLPFLVSNHPLEKFEIOM
KVEWVYVSONSFPESSSDGFGSIRGLINISITVGAKFYOGDIDSTYDLSGFPVS
VYRNPOSTATLVMSPSDKIRGNLSRQAFLLRGSNNYVNSNCELFGHYAMELNGS
SRNINVDVGRKLP"
BASE COUNT 3228 a 2459 c 2065 g 3005 t
ORIGIN
Query Match 99.6%; Score 2803; DB 1; Length 10757;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2814; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 ATGAATGCAATTTTCTGTTAGTGCTCTCGACATNGGATGTTTACTAGTTGT 60
Db 7650 ATGAATGCAATTTTCTGTTAGTGCTCTCTCGACATNGGATGTTTACTAGTTGT 7591
QY 61 TCCACTGTTTTGCTGCAACTGCTGAAATATVAGCCCTCTGATGCTTGAAGCAAGT 120
Db 7590 TCCACTGTTTTGCTGCAACTGCTGAAATATVAGCCCTCTGATGCTTGAAGCAAGT 7531
QY 121 ACTAACACAGCACCCTATCTCTTAAATATGCACTGCAATAGACTATCTCTGACA 180
Db 7530 ACTAACACAGCACCCTATCTCTTAAATATGCACTGCAATAGACTATCTCTGACA 7471
QY 181 GGAGATATTAACCTGCAAAACCTTGGGATTCGCGACCTTAAAGAAAGGTTGTTTCT 240
Db 7470 GGAGATATTAACCTGCAAAACCTTGGGATTCGCGACCTTAAAGAAAGGTTGTTTCT 7411
QY 241 GACACTACGCAATCTTAAAGCTTTCGCGTAAAGGGTACTCATCTTCTTTTAAATAT 300
Db 7410 GACACTACGCAATCTTAAAGCTTTCGCGTAAAGGGTACTCATCTTCTTTTAAATAT 7351
QY 301 AAGCTAGTGTGTAAGGCGACGACTTCTGTTACACTGATAAATCTGTGCTACACA 360
Db 7350 AAGCTAGTGTGTAAGGCGACGACTTCTGTTACACTGATAAATCTGTGCTACACA 7291
QY 361 GGATTTTCGAGTCTTACTTCTTACGCGCCCATCATCGGTAAATCACAAACCCCTCAGA 420
Db 7290 GGATTTTCGAGTCTTACTTCTTACGCGCCCATCATCGGTAAATCACAAACCCCTCAGA 7231
QY 421 AAAGTCGAGTTAAATGTGAGGGGATCTTACATTTGATAAACAATGCACTATTTATTT 480
Db 7230 AAAGTCGAGTTAAATGTGAGGGGATCTTACATTTGATAAACAATGCACTATTTATTT 7171
QY 481 AAACAATTAACGTGAGGAAATGGCGGACCATTTCTACCAAGATCTTCTTGAAA 540
Db 7170 AAACAATTAACGTGAGGAAATGGCGGACCATTTCTACCAAGATCTTCTTGAAA 7111
QY 541 AACAGCAGGGATCGATTTCTTTTGAAGGGAATTAATCGAGCCCAACAGGAAAAAAGT 600
Db 7110 AACAGCAGGGATCGATTTCTTTTGAAGGGAATTAATCGAGCCCAACAGGAAAAAAGT 7051
QY 601 GGGGCTATTGTGCTACTGCTAGTATGATTAACAATAATACGGCTCTTACCCCTTC 660
Db 7050 GGGGCTATTGTGCTACTGCTAGTATGATTAACAATAATACGGCTCTTACCCCTTC 6591
QY 661 TCGACAATAATGTGTAAGCTGAGCGTGAAGCTAATAATGACACAGAACTGTACAAAT 720
Db 6590 TCGACAATAATGTGTAAGCTGAGCGTGAAGCTAATAATGACACAGAACTGTACAAAT 6931
QY 721 ACAAGGAATACGTCCTTGTATTTTCTGAAATATAGTGTGACAGGACCGAGAAATAGA 780
Db 6930 ACAAGGAATACGTCCTTGTATTTTCTGAAATATAGTGTGACAGGACCGAGAAATAGA 6871
QY 781 GGAAGCTCTTCTGAGATGCGGATGTTACATATCTGGGAATCAGAGTGTAACTTCTCA 840
Db 6870 GGAAGCTCTTCTGAGATGCGGATGTTACATATCTGGGAATCAGAGTGTAACTTCTCA 6811
QY 841 GGAACCAACTGTAGCTAATGGCGAGCCATTTATGCTAAGAACTTAACGTGCTTC 900
Db 811 GGAACCAACTGTAGCTAATGGCGAGCCATTTATGCTAAGAACTTAACGTGCTTC 890
```

Db 6810 GGAACCAAGCTAGTAATGCGGAGCCATTATGCTAAGAAAGCTTACACTGGCTCC 6751  
Qy 901 GGGGGGGGGGGGATCTCTTTTCTTACAAATATAGTCCAAAGGTACACACTGCGAGTAT 960  
Db 6750 -GGGGGGGGGGGATCTCTTTTCTTACAAATATAGTCCAAAGGTACACACTGCGAGTAT 6692  
Qy 961 GGTGAGGCAATTTCTATCTAGTGGAGCTGGAGAGTGTAGTCTTTCAGAGAGAGCGGGGAC 1020  
Db 6691 GGTGAGGCAATTTCTATCTAGTGGAGCTGGAGAGTGTAGTCTTTCAGAGAGAGCGGGAC 6632  
Qy 1021 ATTACCTTCAATGGGAATGCCATTGTTGCACTACACCACAACTAGCAAAAAGAAATTC 1080  
Db 6631 ATTACCTTCAATGGGAATGCCATTGTTGCACTACACCACAACTAGCAAAAAGAAATTC 6572  
Qy 1081 ATTACATAGGATCTACTGCAAGATCAGCAATTTAGCTGCAATATCTGGGCAATGATC 1140  
Db 6571 ATTACATAGGATCTACTGCAAGATCAGCAATTTAGCTGCAATATCTGGGCAATGATC 6512  
Qy 1141 TTTTCTACAGATCCGATTTACGTAATACGGCTGCGGATTTCTACAGATCTTAAATCTC 1200  
Db 6511 TTTTCTACAGATCCGATTTACGTAATACGGCTGCGGATTTCTACAGATCTTAAATCTC 6452  
Qy 1201 AATAAGGCTGATGCAAGTAATAGTACAGATTTATGAGTGGTCTGATTTCTTCTGTA 1260  
Db 6451 AATAAGGCTGATGCAAGTAATAGTACAGATTTATGAGTGGTCTGATTTCTTCTGTA 6392  
Qy 1261 AAGCTCTGTAAGTGAAGCAAAAGTTCGACACACTCTACTTACGCTGACAGCTT 1320  
Db 6391 AAGCTCTGTAAGTGAAGCAAAAGTTCGACACACTCTACTTACGCTGACAGCTT 6332  
Qy 1321 GTAACCTTACGTCAGGAAATTTAGTACTTAAAGTGGTCTGCTGATCTGATACGAAAGC 1380  
Db 6331 GTAACCTTACGTCAGGAAATTTAGTACTTAAAGTGGTCTGCTGATCTGATACGAAAGC 6272  
Qy 1381 TTTTACTAGACCGCGGCTCTCTGTTATATGAGTGGGCGACACAGCTTAAAGCAAGT 1440  
Db 6271 TTTTACTAGACCGCGGCTCTCTGTTATATGAGTGGGCGACACAGCTTAAAGCAAGT 6212  
Qy 1441 ACAAGAGGCTCACTTAAACAGTCTTCCATTCCTGTAAGCTCTTTAGGAGGATAG 1500  
Db 6211 ACAAGAGGCTCACTTAAACAGTCTTCCATTCCTGTAAGCTCTTTAGGAGGATAG 6152  
Qy 1501 AAAGTTTAAATGCTGCTGCTGAGCAAGTAAATGTAGCCCTTATGCTGATCTT 1560  
Db 6151 AAAGTTTAAATGCTGCTGCTGAGCAAGTAAATGTAGCCCTTATGCTGATCTT 6092  
Qy 1561 CTTTGTGATTAACCAAGGAATGCTTATGAAATCAGCACTTATGAGAAATCAAGACTT 1620  
Db 6091 CTTTGTGATTAACCAAGGAATGCTTATGAAATCAGCACTTATGAGAAATCAAGACTT 6032  
Qy 1621 TCATTTGTGAGCTCTGCTGCTGAGTACGCAACACTACAGATGTTCCAGCGGTTCT 1680  
Db 6031 TCATTTGTGAGCTCTGCTGCTGAGTACGCAACACTACAGATGTTCCAGCGGTTCT 5972  
Qy 1681 ACACTAGCACTCTTACGCACTATGGTATCAAGTACTTGGGGAATGACTTGGTGTAT 1740  
Db 5971 ACACTAGCACTCTTACGCACTATGGTATCAAGTACTTGGGGAATGACTTGGTGTAT 5912  
Qy 1741 GATACCGCAAGCACTCCAAAGACTAAGACAGCACTAGCTTGGACCAATACAGGCTAC 1800  
Db 5911 GATACCGCAAGCACTCCAAAGACTAAGACAGCACTAGCTTGGACCAATACAGGCTAC 5852  
Qy 1801 CTTCCGAATCTGAGCTGCAAGGACTTATGTTCTTAAATGCTTGGGGAATCTTTTCA 1860  
Db 5851 CTTCCGAATCTGAGCTGCAAGGACTTATGTTCTTAAATGCTTGGGGAATCTTTTCA 5792  
Qy 1861 GACATCCAAAGGATTCAGGATGTCATAGAGAGAGAGTCTTGTGACTTGTGTCAGATCGA 1920  
Db 5791 GACATCCAAAGGATTCAGGATGTCATAGAGAGAGAGTCTTGTGACTTGTGTCAGATCGA 5732  
Qy 1921 GCGTTCTGCGCTGGGAGTGGCCAAATTTCTTATGATTAAGATTAAGAAAGGGAACCGC 1980  
Db 5731 GCGTTCTGCGCTGGGAGTGGCCAAATTTCTTATGATTAAGATTAAGAAAGGGAACCGC 5672

Qy 1981 AAATACGCTAATAATGCTGNGATATGCTATCGAGGTGACAGGCAAACTGTTCTGAA 2040  
Db 5671 AAATACGCTAATAATGCTGNGATATGCTATCGAGGTGACAGGCAAACTGTTCTGAA 5612  
Qy 2041 AACTTAATATGCTTGGCTTGGCAACTCTTGTGTAGCGATAAAGATTTCTTATGCTCT 2100  
Db 5611 AACTTAATATGCTTGGCTTGGCAACTCTTGTGTAGCGATAAAGATTTCTTATGCTCT 5552  
Qy 2101 AAAATATATGCTATACCTATGACAGAGGCTTCTATATCAACACATTTACAGAAATGAT 2160  
Db 5551 AAAATATATGCTATACCTATGACAGAGGCTTCTATATCAACACATTTACAGAAATGAT 5492  
Qy 2161 GGGTTCATAGTGTCTCTTATGATTAACCTTCTGCTTGTGAGTCAATAAACCCCTGCT 2220  
Db 5491 GGGTTCATAGTGTCTCTTATGATTAACCTTCTGCTTGTGAGTCAATAAACCCCTGCT 5432  
Qy 2221 TTAGAAGGCGAGCTCGCTTATAGCCAGTCAATGATGATGTAAGACAAAGTATAGT 2280  
Db 5431 TTAGAAGGCGAGCTCGCTTATAGCCAGTCAATGATGATGTAAGACAAAGTATAGT 5372  
Qy 2281 TATCTGAGTGAAGGTTCTTGGGGAAATATGCTTTTAAATGATGTTGGAGCTTCT 2340  
Db 5371 TATCTGAGTGAAGGTTCTTGGGGAAATATGCTTTTAAATGATGTTGGAGCTTCT 5312  
Qy 2341 TCTCATCTTATCTCGTAATACCTGATGTTGATTAACCTATGCTCCATACATCAAACTG 2400  
Db 5311 TCTCATCTTATCTCGTAATACCTGATGTTGATTAACCTATGCTCCATACATCAAACTG 5252  
Qy 2401 AATCTGACCTTATATGCTCAGACAGCTTCTCGAGAAAGTATACAGAAAGTAACTTTT 2460  
Db 5251 AATCTGACCTTATATGCTCAGACAGCTTCTCGAGAAAGTATACAGAAAGTAACTTTT 5192  
Qy 2461 GATGACAGCAACCTCTTCAATTTATCTTGGCTATAGGGGTGAAGTTGAGAGTCTCT 2520  
Db 5191 GATGACAGCAACCTCTTCAATTTATCTTGGCTATAGGGGTGAAGTTGAGAGTCTCT 5132  
Qy 2521 GATTGTATGCTTCTTATGATCTGACTTATCTATGCTATGCTGATCTTATGCTGCAAT 2580  
Db 5131 GATTGTATGCTTCTTATGATCTGACTTATCTATGCTATGCTGATCTTATGCTGCAAT 5072  
Qy 2581 GATCCCAATGCACTACAGCACTTGTATATCAGCGAGGCTTGTGGAAACTTATGCAAT 2640  
Db 5071 GATCCCAATGCACTACAGCACTTGTATATCAGCGAGGCTTGTGGAAACTTATGCAAT 5012  
Qy 2641 AACTTACAGACAGGCTTCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2700  
Db 5011 AACTTACAGACAGGCTTCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 4952  
Qy 2701 TTTGAAGTGGTGGGCGAGTGTGCTTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2760  
Db 4951 TTTGAAGTGGTGGGCGAGTGTGCTTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 4892  
Qy 2761 GATCTTGGGGTAAAGTTCATTTCTAGAGGCTCTCATGCTCAGAAATTCG 2815  
Db 4891 GATCTTGGGGTAAAGTTCATTTCTAGAGGCTCTCATGCTCAGAAATTCG 4837

RESULT 6  
A81851  
LOCUS A81851 3000 bp DNA PAT 21-JAN-2000  
DEFINITION Sequence 25 from Patent WO9858953.  
ACCESSION A81851  
VERSION A81851.1 GI:6731876  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 3000)  
AUTHORS Madsen, A. and Birke Lund, S.  
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHILAMYDIA PNEUMONIAE  
JOURNAL Patent: WO 9858953-A 25 30-DEC-1998.  
MADSEN ANNA SORIE (DK); BIRKE LUND SVEND (DK)

FEATURES  
source  
CDS

Location/Qualifiers  
1..3000  
/organism="unidentified"  
/db\_xref="taxon:32644"  
259..>3000  
/note="unnamed protein product"  
/protein\_id="CAB69200.1"  
/codon\_start=1  
/db\_xref="GI:6731877"  
/translation="MKSFPMVLVSFLACFTSCSYFAEATENIGPDSFSDSTNIG  
TTPKNTTIDITLNGDITLONIGDASALTKCFSPTEPISFAGYKSLSPNIRK  
SABGALSVTTDKNLSTLGSSTLFLAAPSVTTPBGRKAVKGGDLITDNNCTILF  
KQDCEBNGAISTKSLNLSSTGSLSEBENKSSATGKGAICATGTLITNTATP  
LFSNNIAEAGGAIINSTGCTITGNTSLVSENSVYATAGGALSDADVTISGNS  
VTESSNOAVNGAIAKAKLLASGGGGISFNNTIYQGTAGGGAISLLAEGCSL  
SAEAGDITFNKAIVATTPTQTKRNSIDISGATITLRAISGHSIFEDPIANTPA  
DSDDTLNLKADAGNSNDYSIVSESEKLSDEKAYADNLSTLQOPVLTAGNTYL  
KRGVTLDTKFTPTAGSSVIMDAGTTLKASTEEVTLGLSIPVDSLEKGVYIASA  
ASKNALSGPILLLDNOGNAYENHDLGKTQDFSVOLSLAGTATTDVPAVPVATP  
HYGQGTMGMTWDDTASPTKRTATLAWNTGYLPBPEQGLVPMSLNGSFSDIOA  
IOGVIERSLTLSDRGFMAAGVANFLDKKGEKRYRHRKSGVYIAGAAQCSNLY  
ISFAFQOLFSGSDFLVAKNHTPDYAGAFYIOHTBESGFTIGCLIDLPGSMHKLPL  
LEGOALYSHYNDLKTXYTAPYKSGMGNANMMIAGSSHSVPETLHCFPTAPYI  
KLMLTYTRDPSSEKTEGSRFSDSNLFNLSLTIYKFEFSQDNDSITDLTSLYPS  
LIRNDPKCTTALVIGSASMETYANNLARQALQVRAGSHYAFSPWFEVLGQFVEPDS

BASE COUNT 867 a 597 c 658 g 878 t  
ORIGIN

Query Match 97.4% Score 2742; DB 9; Length 3000;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2742; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGAATCGCAATTTCTCTGTTAGTGTCTCTCTCGACATTCGATGTTTACTAGTGT 60  
DB 259 ATGGAATCGCAATTTCTCTGTTAGTGTCTCTCTCGACATTCGATGTTTACTAGTGT 318  
QY 61 TCCACTGTTTGGTGCACACTGTAATAATAGGCCCTCTGATAGCTTTCAGGGAAGT 120  
DB 319 TCCACTGTTTGGTGCACACTGTAATAATAGGCCCTCTGATAGCTTTCAGGGAAGT 378  
QY 121 ACTAACACAGGACCTTACTCTTAAATAAGACTACTGGAATAGACTATATCTGACA 180  
DB 379 ACTAACACAGGACCTTACTCTTAAATAAGACTACTGGAATAGACTATATCTGACA 438  
QY 181 GGAAATTAATCTGCAAAACCTTGGGATTCGACGCTTAAAGAGGCTGTTTTCT 240  
DB 439 GGAAATTAATCTGCAAAACCTTGGGATTCGACGCTTAAAGAGGCTGTTTTCT 498  
QY 241 GACACTACGGAATCTTAAAGCTTCCGGAAGGGTACTACTCTCTTCTTAAATAT 300  
DB 499 GACACTACGGAATCTTAAAGCTTCCGGAAGGGTACTACTCTCTTCTTAAATAT 558  
QY 301 AAGTCTAGTGTGAGGCGGACACTTCTGTACAACTGATAAAATCTGTGCTAACA 360  
DB 559 AAGTCTAGTGTGAGGCGGACACTTCTGTGTACAACTGATAAAATCTGTGCTAACA 618  
QY 361 GGATTTTCGAGTCTTACTTCTTTAGCGGCCCATCATCGTAACACACCCCTCAGGA 420  
DB 619 GGATTTTCGAGTCTTACTTCTTTAGCGGCCCATCATCGTAACACACCCCTCAGGA 678  
QY 421 AAAGTCTACGTAATGTGAGGGGATCTTACATTTGATAACATGGAACATTTTATTT 480  
DB 679 AAAGTCTACGTAATGTGAGGGGATCTTACATTTGATAACATGGAACATTTTATTT 738  
QY 481 AAACAGATTACTGTGAGAAATGCGGAGCCATTTCTACCAAGAACTCTTCTTGAAA 540  
DB 739 AAACAGATTACTGTGAGAAATGCGGAGCCATTTCTACCAAGAACTCTTCTTGAAA 798  
QY 541 AACAGCAGCGGATTCGATTTCTTTGAAGGAATTAATCGAGCGCAACAGGAAAAAGT 600  
DB 799 AACAGCAGCGGATTCGATTTCTTTGAAGGAATTAATCGAGCGCAACAGGAAAAAGT 858

QY 601 GGGCTATTTTGTCTACTGCTAGTATGATATTAATAATATACGGCTCTACCCCTCTTC 660  
DB 859 GGGCTATTTTGTCTACTGCTAGTATGATATTAATAATATACGGCTCTACCCCTCTTC 918  
QY 661 TCGAACATATTTGCTGAAGCTGCAGGTGAGCTTAAATATGACACAGAACTGATTA 720  
DB 919 TCGAACATATTTGCTGAAGCTGCAGGTGAGCTTAAATATGACACAGAACTGATTA 978  
QY 721 ACAGGAATATGCTCTGTTGTAATTTTTCGAAATATGTGACAGGACCGGAGAAATGA 780  
DB 979 ACAGGAATATGCTCTGTTGTAATTTTTCGAAATATGTGACAGGACCGGAGAAATGA 1038  
QY 781 GGACCTCTTTCGAGATGCGGATGTTACATATCTGGGAATTCAGAGTGAATCTTCTCA 840  
DB 1039 GGACCTCTTTCGAGATGCGGATGTTACATATCTGGGAATTCAGAGTGAATCTTCTCA 1098  
QY 841 GGAACACAGCTGTAGCTAATGCGGAGCCATTTATGCTAAGAACTTACACTGCTTCC 900  
DB 1099 GGAACACAGCTGTAGCTAATGCGGAGCCATTTATGCTAAGAACTTACACTGCTTCC 1158  
QY 901 GGGGGGGGGGGGATCTCTCTTTCTAACAATATGTCACAAAGGTACCACTGCAGTAAT 960  
DB 1159 GGGGGGGGGGGGATCTCTCTTTCTAACAATATGTCACAAAGGTACCACTGCAGTAAT 1218  
QY 961 GGTGAGCCATTTCTATACCTGCGAGCTGAGAGTGTAGTCTTTCAGCAGAGAGGGAC 1020  
DB 1219 GGTGAGCCATTTCTATACCTGCGAGCTGAGAGTGTAGTCTTTCAGCAGAGAGGGAC 1278  
QY 1021 ATTAACCTCAATGGAATGCCATGTTTGCAATACACCCACAACTACAAAAAATTTCT 1080  
DB 1279 ATTAACCTCAATGGAATGCCATGTTTGCAATACACCCACAACTACAAAAAATTTCT 1338  
QY 1081 ATTGACATGAGATCTACGCAAGATCAGCAATTTAGTGCATATTCGGGCAATGATC 1140  
DB 1339 ATTGACATGAGATCTACGCAAGATCAGCAATTTAGTGCATATTCGGGCAATGATC 1398  
QY 1141 TTTTCTACGATCCGATTAATGCTAATACGCTGCGGATTTCTACAGATCTTAAATCTC 1200  
DB 1399 TTTTCTACGATCCGATTAATGCTAATACGCTGCGGATTTCTACAGATCTTAAATCTC 1458  
QY 1201 AATAAGCTGATGCAAGTAAATATGATAGATTAATGAGGCTGATGTTTTTCTGCTGAA 1260  
DB 1459 AATAAGCTGATGCAAGTAAATATGATAGATTAATGAGGCTGATGTTTTTCTGCTGAA 1518  
QY 1261 AAGCTCTGAGATGAGCAAAAGTTGCAGACAACTCTTACGCGAAGCAGCT 1320  
DB 1519 AAGCTCTGAGATGAGCAAAAGTTGCAGACAACTCTTACGCGAAGCAGCT 1578  
QY 1321 GTAACCTAATCGCAGAAATTTAGTACTTAAACGTGTGTCACTCGATACGAAAGGC 1380  
DB 1579 GTAACCTAATCGCAGAAATTTAGTACTTAAACGTGTGTCACTCGATACGAAAGGC 1638  
QY 1381 TTTACTAGACCGGGTCTCTGTTATTAATGATGCGGGCACACGTTAAAGCAAGT 1440  
DB 1639 TTTACTAGACCGGGTCTCTGTTATTAATGATGATCGGGCACACGTTAAAGCAAGT 1698  
QY 1441 ACAGAGAGGCTAATTAACAGGCTTTCATCTGTAGACCTTTTAAAGGAGGATG 1500  
DB 1699 ACAGAGAGGCTAATTAACAGGCTTTCATCTGTAGACCTTTTAAAGGAGGATG 1758  
QY 1501 AAAGTCTAATTTGCTCTGCTGACAGCAAGTAAATATGAGCCCTTACTGCTCCGATTTCT 1560  
DB 1759 AAAGTCTAATTTGCTCTGCTGACAGCAAGTAAATATGAGCCCTTACTGCTCCGATTTCT 1818  
QY 1561 CTTTGTGATACCAAGGAATGTTATGAAATATCAGACTTAAAGAAACTCAAGACTTT 1620  
DB 1819 CTTTGTGATACCAAGGAATGTTATGAAATATCAGACTTAAAGAAACTCAAGACTTT 1878  
QY 1621 TCATTGTGAGGCTCTGCTGCTGAGTACGCAACACTACAGATGTTTCAGAGGCTTCT 1680  
DB 1879 TCATTGTGAGGCTCTGCTGCTGAGTACGCAACACTACAGATGTTTCAGAGGCTTCT 1938

QY 1681 ACAGTAGCACTCTACGACTATGGATGATCAGGTAAGTCTGGGAATGACTTGGTTGAT 1740  
 |||||||  
 Db 1939 ACAATGACACTCTTACGACTATGGATGATCAGGTAAGTCTGGGAATGACTTGGTTGAT 1998  
 QY 1741 GATACCGCAACACTCCAAAGACTAGACAGCAGCATTTAGCTTGGACCAATACAGGCTAC 1800  
 |||||||  
 Db 1999 GATACCGCAACACTCCAAAGACTAGACAGCAGCATTTAGCTTGGACCAATACAGGCTAC 2058  
 QY 1801 CTTCGGAATCTGAGCTGCAAGGACCTTTAGTCTTAATAGCTTTGGGATCTTTTCA 1860  
 |||||||  
 Db 2059 CTTCGGAATCTGAGCTGCAAGGACCTTTAGTCTTAATAGCTTTGGGATCTTTTCA 2118  
 QY 1861 GACATCCAAAGGATTCAGGATGATAGAGAAAGTCTTTGACTCTTTGTCAGATCGA 1920  
 |||||||  
 Db 2119 GACATCCAAAGGATTCAGGATGATAGAGAAAGTCTTTGACTCTTTGTCAGATCGA 2178  
 QY 1921 GCGTCTGGGCTGGGAGTCCCAATTTCTTAGATAAAGTAAGAAAGGGGAAAAACGC 1980  
 |||||||  
 Db 2179 GCGTCTGGGCTGGGAGTCCCAATTTCTTAGATAAAGTAAGAAAGGGGAAAAACGC 2238  
 QY 1981 AAATACCGTATTAATCTGGTGGATGATGATGAGAGGAGGCAAACTGTTCTGAA 2040  
 |||||||  
 Db 2239 AAATACCGTATTAATCTGGTGGATGATGATGAGAGGAGGCAAACTGTTCTGAA 2298  
 QY 2041 AACTTAATGACTTGGCTTTGGCAACTCTTTGGTAGCGATTAAGATTTCTTAGTGCCT 2100  
 |||||||  
 Db 2299 AACTTAATGACTTGGCTTTGGCAACTCTTTGGTAGCGATTAAGATTTCTTAGTGCCT 2358  
 QY 2101 AAAATATACATGATATCTGATGAGAGGCTTCTATATCCAAACATTTAGAGATGATG 2160  
 |||||||  
 Db 2359 AAAATATACATGATATCTGATGAGAGGCTTCTATATCCAAACATTTAGAGATGATG 2418  
 QY 2161 GGGTCTAGTGTCTCTTATAGATAAATCTTGGCTCTTGGAGTCAATAACCCCTGTT 2220  
 |||||||  
 Db 2419 GGGTCTAGTGTCTCTTATAGATAAATCTTGGCTCTTGGAGTCAATAACCCCTGTT 2478  
 QY 2221 TTAGAAGGAGGAGCTCGCTTTATAGCCACGTCAGTAAATGATCGAAGTAATAGTGC 2280  
 |||||||  
 Db 2479 TTAGAAGGAGGAGCTCGCTTTATAGCCACGTCAGTAAATGATCGAAGTAATAGTGC 2538  
 QY 2281 TATCCTGAGGTGAAGGCTTGGGGGAATATGCTTTTAACATGATTTGGAGCTTCT 2340  
 |||||||  
 Db 2539 TATCCTGAGGTGAAGGCTTGGGGGAATATGCTTTTAACATGATTTGGAGCTTCT 2598  
 QY 2341 TCTCATCTTATCTGTAATACCTGATCTTTTGTATACCTATGCTCCATACATCAAACTG 2400  
 |||||||  
 Db 2599 TCTCATCTTATCTGTAATACCTGATCTTTTGTATACCTATGCTCCATACATCAAACTG 2658  
 QY 2401 AATCTGACATATATAGCTGAGGACAGCTTCTCGGAGAAAGGTACAGAAAGAAATCTTTT 2460  
 |||||||  
 Db 2659 AATCTGACATATATAGCTGAGGACAGCTTCTCGGAGAAAGGTACAGAAAGAAATCTTTT 2718  
 QY 2461 GATGACAGCAACTCTTCAATTTATCTTGTCCATATAGGGGTGAAGTTTGAAAGTCTCT 2520  
 |||||||  
 Db 2719 GATGACAGCAACTCTTCAATTTATCTTGTCCATATAGGGGTGAAGTTTGAAAGTCTCT 2778  
 QY 2521 GATTTGTAATACCTTTCTTATGATCTGATCTTATCCATATGCTCTGATCTTATCCGCAAT 2580  
 |||||||  
 Db 2779 GATTTGTAATACCTTTCTTATGATCTGATCTTATCCATATGCTCTGATCTTATCCGCAAT 2838  
 QY 2581 GATCCCAAAAGGACTAGCACTTGTATACAGGAGCCTCTTGGGAAACCTTATGCCAAT 2640  
 |||||||  
 Db 2839 GATCCCAAAAGGACTAGCACTTGTATACAGGAGCCTCTTGGGAAACCTTATGCCAAT 2898  
 QY 2641 AACTTACGACAGAGGCTTGGCAAGTGTGACAGGAGCAGTACAGGCTTCTCTATG 2700  
 |||||||  
 Db 2899 AACTTACGACAGAGGCTTGGCAAGTGTGACAGGAGCAGTACAGGCTTCTCTATG 2958  
 QY 2701 TTGGAAGTGTGGGCTGCTTGTGAAAGTCTGATGATCC 2742  
 |||||||  
 Db 2959 TTGGAAGTGTGGGCTGCTTGTGAAAGTCTGATGATCC 3000

## RESULT 7

CPOMP54/c 6030 bp DNA BCT 22-JAN-1999  
 LOCUS Chlamydia pneumoniae omp5 and omp4 genes.  
 DEFINITION AJ001311  
 ACCESSION AJ001311.1 GI:3255934  
 VERSION 1  
 KEYWORDS omp4 gene; omp5 gene; outer membrane protein 4; outer membrane protein 5.

SOURCE Chlamydia pneumoniae.

ORGANISM Chlamydia pneumoniae.

REFERENCE Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

AUTHORS 1 (bases 1 to 6030)

TITLE Knudsen, K.

JOURNAL Direct Submission

Submitted (29-AUG-1997) Knudsen K., Department of Medical Microbiology and Immunology, University of Aarhus, The Bartholin building, University of Aarhus, DK-8000 Aarhus C, DENMARK

2 (bases 1 to 6030)

REFERENCE Knudsen, K., Madsen, A.S., Mygind, P., Christensen, G. and Birkelund, S.

Identification of two novel genes encoding 97- to 99-kilodalton outer membrane proteins of Chlamydia pneumoniae

Infect. Immun. 67 (1), 375-383 (1999)

99081766

FEATURES

source

location/Qualifiers

1..6030

/organism="Chlamydia pneumoniae"

/isolate="CDC/CWL-029/VR-1310"

/db\_xref="taxon:83558"

/clone="pEX2-18"

/clone="B8F3"

/clone\_11b="pbluescript"

/dev\_stage="elementary body"

/complement(1..2804)

/gene="omp5"

/complement(<1..2742)

/gene="omp5"

/note="putative lipoprotein"

/codon\_start=1

/product="outer membrane protein 5"

/protein\_id="CA04671.1"

/db\_xref="GI:3255935"

/db\_xref="SPTREMBL:086163"

/translation="MKSQPSWLVLSSTLACFTSCSTVFATATNIGSPDSFDGNTG  
TYPKMTTGIDITLGLDITLQMLGDSALTKCFSDTESLSFAGYSLSPNTKS  
SABGALSVYTDKNSLTGFSLSFLAAPSIVTTPSGKAVCGGDLTFDNNGTILF  
KQYCEENGGAISTKMLSLKNSGTSISPEGNKSATGKGATCATGTVDTNNVAPT  
LBSNNIAEAGAIINSTGCTTGNTSIVPESSVTRTANGGALSADAVTISGNS  
VFSGQAYANGAATYAKKTLTLASGGGCGISFSNNIVQTTAGNGAISTLAAGCSL  
SASAGDITFNGNAIVATTPQTKRNSIDIGSTAKITNLRAISGHSIFPYDPTAGNTVA  
DSYDTLNKADGNSSTDYSGSIVFSGEKISEDEAVADNLTSYLKQPYVLTAGNTVA  
KRGVTLDTKGTFTAGSSVLMADGTLTKASTEVEVTLGLSIPVDSLGEKKVIAASV  
ASKNVLSGFIILLDNOGNAYENHDGKQDFEVSLSAGTATTTDVPVAVVAPT  
HYGQGTGWTWDTPTASTPKTKATLATLNTGNYLNPENOCGLVPSNLSMGSRSDIOA  
IGQVIRSAITLCSDBGFPAAGYANFLDNDKREKKRYKRSQYVIAAGAOPTCSNLI  
ISFAFCQLFSGSDKDLVANKHDTYGAFTLQHTPCSGITGLDLKLPESWSHAPYI  
LEQOLAVSHVSDNLKRTYAYPEVKSQWGNANMMLGASSHSPYELHCFDPAVYI  
KLNTLYIRDSFSEKGTGERSFEDSNLFINLSLPIGVKFEKFSQCNDFSYDLTLYVAPD  
LHNDKCTTALVLSGASWETVANNIARQALGVBAQSHVAFSPMFEVLQFVEVNGS  
"

mat\_peptide complement(1..2697)

/gene="omp5" /product="outer membrane protein 5"

sig\_peptide complement(2698..2742)

/gene="omp5" complement(2747..2752)

RBS /gene="omp5"

/note="putative" /note="putative"

-10..signal complement(2777..2782)

/gene="omp5" /note="putative"







D	b	1362	TTTACTGAGACCGGGGGTTTCTCTGTTATTATGATGCGCGGACACAAGTTAAAGCAAGT	1303
O	y	1441	ACAGAGGAGGTCACTTTTAAACAGGTCTTTTCCATTCTGTAGACTCTTTAGCGGAGGGTAAG	1500
D	b	1302	ACAGAGGAGGTCACTTTTAAACAGGTCTTTTCCATTCTGTAGACTCTTTAGCGGAGGGTAAG	1243
O	y	1501	AAATTTGTAATTCGTGCTTTTGAGACAAAGTAAAAATGTAGCCCTTAGTGTCCGATTCTT	1560
D	b	1242	AAATTTGTAATTCGTGCTTTTGAGACAAAGTAAAAATGTAGCCCTTAGTGTCCGATTCTT	1183
O	y	1561	CTTTTGGATTAACCAAGGAATGCTTTATGAAAAATCAAGACTTATGAAAAATCAAGACTTT	1620
D	b	1182	CTTTTGGATTAACCAAGGAATGCTTTATGAAAAATCAAGACTTATGAAAAATCAAGACTTT	1123
O	y	1621	TCATTTTGTGAGCTCTCTGCTCTGGGTACTGCAACAATCACAGATGTTCCAGCGGTCTCT	1680
D	b	1122	TCATTTTGTGAGCTCTCTGCTCTGGGTACTGCAACAATCACAGATGTTCCAGCGGTCTCT	1063
O	y	1681	ACAGTACCAACTCTCTACGCACTATGGGTATCAAGTACTTGGGGAAATACCTTGGGTGAT	1740
D	b	1062	ACAGTACCAACTCTCTACGCACTATGGGTATCAAGTACTTGGGGAAATACCTTGGGTGAT	1003
O	y	1741	GATACCGCAAGCACTCCAAAAGACTAGACAGGACATTAGCTTGGACCAATACAGGGCTAC	1800
D	b	1002	GATACCGCAAGCACTCCAAAAGACTAGACAGGACATTAGCTTGGACCAATACAGGGCTAC	943
O	y	1801	CTTCCGAACTCTGAGGCTCAAGGACCTTTTATGTTCTTAATAGCTTTGGGAGATCTTTTCA	1860
D	b	942	CTTCCGAACTCTGAGGCTCAAGGACCTTTTATGTTCTTAATAGCTTTGGGAGATCTTTTCA	883
O	y	1861	GACATCCACAGCATTCAGAGTGTCAATAGAGAAAGTGGTTTACTCTTTGTCAGATCGA	1920
D	b	882	GACATCCACAGCATTCAGAGTGTCAATAGAGAAAGTGGTTTACTCTTTGTCAGATCGA	823
O	y	1921	GGCTTTGGGGCTGGGAGTGGCGGCATTTTCTTAGATTAAGATTAAGAAAGGGGAAAAAGC	1980
D	b	822	GGCTTTGGGGCTGGGAGTGGCGGCATTTTCTTAGATTAAGATTAAGAAAGGGGAAAAAGC	763
O	y	1981	AAATACCGTATAAATCTGGTGGATATGCTATCGGAGGTGACGCCAAACTTGTTCGAA	2040
D	b	762	AAATACCGTATAAATCTGGTGGATATGCTATCGGAGGTGACGCCAAACTTGTTCGAA	703
O	y	2041	AACTTAATTAGCTTTGCTTTTGCCAACTCTTTGGTAGCGATTAAGATTTCTTAGTGGCT	2100
D	b	702	AACTTAATTAGCTTTGCTTTTGCCAACTCTTTGGTAGCGATTAAGATTTCTTAGTGGCT	643
O	y	2101	AAAAATCATCTATATACATATGAGGAGGCTCTTATATCCAAACATATACAGAAATGATG	2160
D	b	642	AAAAATCATCTATATACATATGAGGAGGCTCTTATATCCAAACATATACAGAAATGATG	583
O	y	2161	GGGTTCATAGGTTGTCTCTTAGATAAATCTCTGGCTCTGGAGTCATAAAACCCCTGGT	2220
D	b	582	GGGTTCATAGGTTGTCTCTTAGATAAATCTCTGGCTCTGGAGTCATAAAACCCCTGGT	523
O	y	2221	TTTGAAGGCGAGCTGCTTTATAGCCACGTCAGTAAATGATCTGAAACAAATATACTGCG	2280
D	b	522	TTTGAAGGCGAGCTGCTTTATAGCCACGTCAGTAAATGATCTGAAACAAATATACTGCG	463
O	y	2281	TATTCCTGAGTGAAGAGTTCTTTGGGGGAATATGCTTTTAACATGATGTTGGAGCTTCT	2340
D	b	462	TATTCCTGAGTGAAGAGTTCTTTGGGGGAATATGCTTTTAACATGATGTTGGAGCTTCT	403
O	y	2341	TTCATTTCTTATCTCTGAATACCTGCATTTGTTTGTATGACTATGCTCCATACATCAACTG	2400
D	b	402	TTCATTTCTTATCTCTGAATACCTGCATTTGTTTGTATGACTATGCTCCATACATCAACTG	343
O	y	2401	AATTCGACCTATATACGTGAGACAGCACTTCTGGAGAAAGGTACAGAAAGAAATCTTTT	2460
D	b	342	AATTCGACCTATATACGTGAGACAGCACTTCTGGAGAAAGGTACAGAAAGAAATCTTTT	283
O	y	2461	GATGACAGCAACTCTTTCATTTATCTTTTGCTATAGGGGTGAAGTTTGAAGAATTCTCT	2520
D	b	282	GATGACAGCAACTCTTTCATTTATCTTTTGCTATAGGGGTGAAGTTTGAAGAATTCTCT	223

QY	2501	GATTGTAATACACTTTTCTTTATGATGACGCTTATCTTAAAGTTCGAGATCTATACCGCAAT	2580
QY	2551		
Db	222	GATTGTATATGACTTTTCTTTATGATGACGCTTATCTTAAAGTTCGAGATCTATACCGCAAT	163
QY	2581	GATCCCAATATGACACTATACAGCACTTGTAATACAGCGGAGCCCTTGGGAAACTTATGCCAAT	2640
Db	162	GATCCCAATATGCACTATACAGCACTTGTAATACAGCGGAGCCCTTGGGAAACTTATGCCAAT	103
QY	2641	AACATTACACGACAGCGCTTTCGCAAGTGCGCTGCAGCGACTACTACGCTTCTCTCTATG	2700
Db	102	AACATTACACGACAGCGCTTTCGCAAGTGCGCTGCAGCGACTACTACGCTTCTCTCTATG	43
QY	2701	TTTGAAGTCTCGGCGCACTTTCCTTTTGAAGTTCGATGATCC	2742
Db	42	TTTTGAAGTCTCGGCGCACTTTCCTTTTGAAGTTCGATGATCC	1

```

RESULT      8
AB1837
LOCUS       AB1837
DEFINITION  AB1837
ACCESSION   AB1837
VERSION      AB1837.1
KEYWORDS     GI:6731869
SOURCE       unidentified.
ORGANISM     unidentified
REFERENCE    1 (bases 1 to 2757)
AUTHORS      Madsen,A. and Birkelund,S.
TITLE        NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL      Patent: WO 9858953-A 11 30-DEC-1998;
             MADSEN ANNA SOPHE (DK); BIRKELUND SVEND (DK)
FEATURES
SOURCE       1..2757
             /organism="unidentified"
             /db_xref="taxon:32644"
BASE COUNT   770 a 564 c 642 g 781 t
ORIGIN

```

Query Match	23.8%	Score 670.6	DB 9	Length 2757
Best Local Similarity	55.8%	Pred. No.3.2e-190		
Matches 1570	Conservative	0	Mismatches 1159	Indels 84
				Gaps 12
QY	1	ATGAATTCGCAATTTTCCGTGTAGTGCCTCTTCGACATTGGCATGTGTTTACTAGTTGT	60	
Db	1	ATGAGATGTCCTCTTTTCCCTGTTATTATATATCTTATCTAGCCCTTTCCTCTTATAG	60	
QY	61	TCCACTGTTTTTGGTCGCAACTGCTGAAAATATAGGCCCTCTGATAGCTTTGACGGAGT	120	
Db	61	AGTGTTCGCGAGATGCTGCCGATCTCCATTTAGGAGATCTGACAGTTAATAGTAT	120	
QY	121	ACTAACACAGGCAOCTATATCTCTTAAATAATAGACTAC-----TGSATATAGCTAT	171	
Db	121	ACAAACACACACAGAAATTTACTTCCTTAACCGGCACATCTTGATGCTAGTGGCGACAGCTAT	180	
QY	172	ACTCTGACAGAGATATTAACCTCTGCCAAAACCTTGGGGATTGGCGACGTTTAACGAAGGAT	231	
Db	181	ATTCCTGATGGGAGTCTCGATTAAGCCAAACAGGAGGAAACAAACAGCTTAAACCAAGT	240	
QY	232	TGTTTTTCTGACACAGCGAATCTTTAAGCTTTGGCGGTAAAGGGTAACTCACTTCTTTT	291	
Db	241	TGTTTTTCTTAACACGACGAGAAATCTTTACCTCTTTAGGAAACGGATTTTCTCTCATTTT	300	
QY	292	TTAAATATTAAAGTCAGAGCTGAAGGCGACAGCACTTTCTGTTAACATGATPAAATAATG	351	
Db	301	GACAATATTATTTCGTACT--GTTCAGAGTGTGTTGTTAGAAATAACGACAGCTTCT	357	
QY	352	TCGCTAACAGGATTTTCGAGTCTTACTTTCTTAGCGGCCCATCATCGGTATCAACACC	411	
Db	358	GGGATTTAGAAATCTCAGGATTTTCAACTCTCGGATGCTGGACGCTCCTAGG-----	412	





gene complement(15438..16376)  
 /gene="xerc"  
 /note="CPH0024"  
 CDS complement(15438..16376)  
 /gene="xerc"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Integrase/recombinase"  
 /protein\_id="AADI8177.1"  
 /db\_xref="GI:4376276"  
 /translation="MIASISFLDYIKVKSASPHLRNYCLDLNGLKIFLEERGLA  
 PSSPOLATKRVSELPSLPFLKEHRYVIALINGRAKRTIKCLSSIKSFAYC  
 VIOKILLENPAETIHGPRLEKELSPMLAYOVELATPDISKHLDRLCMLELYS  
 SGLRISEIVAVNKODELSTHLIRIGKKEKRIIIVTSNAIQMIDLYNHDPKRLIE  
 KDPQAFILNRFGRISIRSIDRSFOEYLSKQTHIHTIIRHTIATIMLESMDLK  
 TIOALIGHSLEHTTYTVOVSXKLKQTHQEAHPHA"

BASE COUNT 4710 a 3330 c 3665 g 4743 t  
 ORIGIN

Query Match 23 68; Score 663; DB 1; Length 16448;  
 Best Local Similarity 55.68; Pred. No. 2.3e-148;  
 Matches 1565; Conservative 0; Mismatches 1165; Indels 83; Gaps 12;

QY 1 ATGAAATCGCAATTTCTGCTAGTCTCTCTGACATTTGGCATTTTACTAGTTGT 60  
 DB 3196 ATGAGATCGCTTTTCTGCTAGTCTCTCTGACATTTCTGCTCTCTTATG 3255  
 QY 61 TCCACTGTTTTGCTGCACTGCTGAAAATATAGCCCTCTGATAGCTTTGACGGAAT 120  
 DB 3256 AGTGTTCACATATGCTGCGATCTCACATTTAGGAGTGTGACAGTTAATGTGAT 3315  
 QY 121 ACTAAGACAGGACGCTTATCCATAAATAGCACTAC-----TGAATATAGACTAT 171  
 DB 3316 ACAAGCACACAGATTTACTCTTAAGCGGCACTTCTGATGCTAGTGCACGACCTAT 3375  
 QY 172 ACTGTGACAGAGATATATACCTTGCAAAACCTTGGGAGTTGCGGAGCTTTAAGCAAGGT 231  
 DB 3376 ATCTCGATCGGAGTGTCTGATATAGCCAAAGGGAACAAACGACCTTAACACAGT 3435  
 QY 232 TGTTTTCTGACACATCGCAATCTTTAAGCTTTGCCGTAAGGGGTACTACTTTCTTT 291  
 DB 3436 TGTTTTCTGACACATCGCAAGAACTTTACCTTTAGGGAACGATTTCTCTTCATTTT 3495  
 QY 292 TTTAAATTTAAGTCTAGTCTGTAAGGCGGCACTTCTGTTACAACTGATATAAATCTG 351  
 DB 3496 GACAATATTTTTCGTACT---GTTGCAAGGTTTGTGTTAGCAATTAAGCAAGCTTCT 3552  
 QY 352 TGCCTAAGAGATTTTGCAGTCTTACTTTTAGCGGCCCATCATCGGTATTCACAAC 411  
 DB 3553 GGGATTTACGAAATTTCTAGGATTTTCACTTCGATGCTTGCAGCTCCTAGGA----- 3607  
 QY 412 CCTCAGGAAAGGTGAGTTAATGTGAGGGGATCTTACATTTGATTAACATGAACT 471  
 DB 3608 -CCACAGGTAAAGAGCATTAATAATTCGATGCTGTGTTGGAATTAAGGAAT 3666  
 QY 472 ATTTTATTAAGATTTCTGTAAGAAATGCGCGACCATTTTACCAAAATCTT 531  
 DB 3667 CTGTGATCTTAATGAAATGCTCTAGTAAATGCGGGAACCATCATTAAGAACTTTG 3726  
 QY 532 TCTTTGAAAAACAGACGGGATGATTTCTTTTGAAGGAATAAATGAGCGCAACGAGG 591  
 DB 3727 TCTTTGACTGGAGTACGGGGTTTGTAGCTTCTGCAATAGCTGTC-----G 3777  
 QY 592 AAAAAAGGTGGGCTATTTGTCTAGTCTGATGATTTTACAAATATATAGGCTCT 651  
 DB 3778 CAACAAGGGGAGCGATCTATGCTTGTGAGCTGTGATTTCTGGAATGAGAGATC 3837  
 QY 652 ACCCTCTTCGCAATATTTGCTGAAGCTGCAAGGTGAGGACCTTAATATACAGGAAC 711  
 DB 3838 TTGAGCTTGGGAACAAGATGCGACACATCTAGGAGCGCGATCTCTGTAAGGAAC 3897  
 QY 712 TGTACAAATTACAGGAATAGCTCTGTATTTTCTGAAAAATAGTGTGACAGGACCGCA 771

DB 3898 CTGTGATCTCCAAATACCAAAATATCTTTTGCATGCTGCAAAAGCACTACAAATGCG 3957  
 QY 772 GGA-----AATGAGAGGCTCTTTGGAATGCCATGTTTACCATATCTGGG 819  
 DB 3958 GGACATATGATATGTAACAAAGCGAGGGCGCAACCCAACTTATCTTGTGTCAGGA 4017  
 QY 820 AATCAGAGTGAATCTTCTCAGAAACAGCTGACTAATGCGGAGGACATTTATGCT 879  
 DB 4018 AATGAGAGCTGCAATTTTCTGATTAACACAGCAGGAATATGTGAGGTGGCATTTATACC 4077  
 QY 880 AAGAAGCTTACACTGCTCCGGGGGGGGGGGGGATTCCTTTTCTAAATATAGTC 939  
 DB 4078 AAAAAATGGTGTATCTCAGACGAGGAGAGGTATTAATTTCTAACAAACAAAGCTGCG 4137  
 QY 940 CAAGTACCACTGAGGTAATGATGAGCCATTTCTATCTGCGAGCTGAGAGGTAGT 999  
 DB 4138 AATGCTACTCC-----TAAAGAGGGGCAATTCGATTTCTGATGCTGAGAGATTAGC 4191  
 QY 1000 CTTTACAGAGAGGAGGACATTAACCTTCAATGGAATGCCATTTGTCAACT---ACA 1056  
 DB 4192 ATTTCTCAATCTCGGCAATATCATTTTGCAGGCAATTAACGACACTACAGGAAT 4251  
 QY 1057 CCACAACTACAAAAGAAATTTATGACATAGAGATCTACTGCAAGATCAGGAATTTA 1116  
 DB 4252 CCTGCGAGTGTGACAGAAATGCTATGATCTTGATGCAATGCAAAATTTTAAATCTC 4311  
 QY 1117 CGTCAATATCTGGGCTATGATCTTTTCTAGATCCGATTAATGAGGCTGCG 1176  
 DB 4312 CGAGCGAGCTGGGGAATTAAGTATTTTCTATGATC-----CTATACAGAGCTCA 4362  
 QY 1177 GATTCCTACATTAATTAATTCATTAAGGATGATGAGTATTAATGATATAGT 1236  
 DB 4363 GAGCTACTGATAGCTCTCTTTGAATTAAGCTGACGAGATCTGAAATTAATCTAGGA 4422  
 QY 1237 GGGTCAATGTTTTTCTGCTGAAAAGCTCTCTGAGATGAACAAAGTTGCACACAC 1296  
 DB 4423 GGGTCAATGTTTTTCTGCTGAAAAGCTCTGAGATGAACAAAGTTGAACCTGACAAAT 4482  
 QY 1297 CTGACTCTACGCTGGAAGGAGGCTGTAACCTTAACGCAAGAAATTAATGATTAACGT 1356  
 DB 4483 CTGAAGCTTACATTAACAGGCTGTAGAGCTGTGACAGTGTCTTAATGAAAGAT 4542  
 QY 1357 GGTGTCACTGATGATGAAGAGGCTTACTCAGACCGGGGTTCTCTGTTATATGAT 1416  
 DB 4543 GAGTGACTGATGATGAATTAACAGGCTGTAGAGCTGTGACAGTGTCTTAATGAAAGAT 4602  
 QY 1417 GCGGCAACAGTTAAAGCAAGTACAGAGAGGCTCACTTAACAGGCTTTTCCATCTCT 1476  
 DB 4603 GGAAGGACTACTTTTGAAGGCAAGGCTGAGGGGCTCACTCAATGAGGCTTACCAATTAAT 4662  
 QY 1477 GTAGACTTTTAGCGGAGGGAAGAAAGTTGTAATCTCTCTGAGCAAGTAAATAT 1536  
 DB 4663 ATGATTTCTTAGGTGGAACAATTAAGCTATCTTAATGAGCGAGGCAAGTAATGAGAT 4722  
 QY 1537 GTAGCCCTTATGCTGATCTTCTTTTGGATTAACCAAGGAATGCTTTAATAATCAC 1596  
 DB 4723 GTTGCTTATCAGGGCTTATCATGCTTGTAGATGCTCAGGGAACATTTATGAGCATCAT 4782  
 QY 1597 GACTTGAAGAAACCTCAAGACTTTTCAATTTGTGCAAGCTCTGCTGTGCTACTGCAACA 1656  
 DB 4783 AATCTCAGTCAACAGCAGGCTTCTTCTTAATAGAGCTTCTCAGCAAGGAAGATGACT 4842  
 QY 1657 ACTAAGATTTCTCAGGAGGTTCTCAGTGAAGCAACCTCAGCAATTTGGGTATCAAGT 1716  
 DB 4843 ACTACATATATCCCGATACCCCAATTTCAATTAATCAATCAATTAATGATTAAGGAG 4902  
 QY 1717 ACTTGGGAATGATGCTGCTGATGATACGCAAGCACTCCAAAGACTAAGACAGCACA 1776  
 DB 4903 AACTGGAATATTTGTTGGTGC-----ACGATGCAACTGCAAAAAACAAAAATGCTAC 4957  
 QY 1777 TTAGCTTGAACCAATACAGGCTACTCTGCAATCTGAGCTGCAAGGACTTTAGTTCT 1836

```

Db 4958 TTAACCTTGAGATAAACAAGATACAGCCGCAATCCAGAACGTCAGAGCACTTGGTTCTT 5017
Oy 1837 AATAGCCCTTGGGATGCTTTTTCAGACATCCAAAGCATCAAGTGCATATGAGAGAACT 1896
Db 5018 AATAGCCCTTGGGATGCTTTTTCAGATGTCGCGTCCGATTCAGACCTCATATGACCGGAGC 5077
Oy 1897 GCTTTCATCTTTTTCAGATCGAGGCTTTCGCGCTGCGGATCGGCAATTTCTTATGAT 1956
Db 5078 ACAAGTTCGTTATCTTCGTCAACAATTTTGTGGATATCAGAAATCCGCGATTTTTCAT 5137
Oy 1957 AAGATAGAAGAGGGGAAAAAGCAAAATACCGCATATAATCTGTGTGATATGATACGGA 2016
Db 5138 GAAATAGAGAAAGGAAACCAAGTAGTATCTACGCGGGGTATGACATTAAGA 5197
Oy 2017 GGTCCACGCGCAAACTTGTCTGAAACCTTAATAGCTTGGCTTTGGCCACTTTCGT 2076
Db 5198 GGAGGATCTTTCACGCGCTTCGAAATTTCTTAAATTTTGTCTTTTGTGACACTTTTGGC 5257
Oy 2077 AGCATATAAGATTTCTTAGTCGCTAAAAATCATCTGATACCTATGACGAGGACCTTCTAT 2136
Db 5258 TACGACAAAGGACCATCTTGTGGCTAAGACCATACCATGATATGACGAGGCAATGAGT 5317
Oy 2137 ATCCAAACATTTACAGATGATGAGTGTAGGCTTCTCTTATAGATAAACTTCTGCGC 2196
Db 5318 TACGACACCTCGGAGAGTCTA-----AGACCCCTCGCTTAAGATTTGTGCAAGA 5365
Oy 2197 TCTTGGAGTCTAAACCCCTGCTTTAGAAAGGACGCTCGCTTATAGCAAGTACATAT 2256
Db 5366 AATTCGACTCCCTACCTTTTGTCTTCAATGCTGGGTTTGTCTATGAGCATACGACAT 5425
Oy 2257 GATCTGAAGACATTTACTGCTATCCTGAGGTGAAGGATCTTGGGGAATATGCT 2316
Db 5426 AACTGACACCAAAAGTACACGCTATCTCTTTAAAGGAACTGGGAAATGATGCC 5485
Oy 2317 TTTAATCATGATGTTGGAGCTTCTTCATTTCTATCT--GAATACCTGCAATGTTTT 2373
Db 5486 TTTGCTATGATGATGAGGAGGATATCCCGGTAGTCTTCAGAGACGCGCTTGGGTG 5545
Oy 2374 GATACATGCTCATACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2433
Db 5546 GATACCCACACGCAATTTTCAACCTAGAGATGATGATGATGATGATGATGATGATGATGAT 5605
Oy 2434 GAGAAGGATACAGAGAAAGATCTTTGATGACAGCAACCTTCAATTTATCTTTGCT 2493
Db 5606 GAAAGCGGACAGAGAGCGCTTCTTCCAAAGTAAAGACCTTTCATCTAGCGTTTCT 5665
Oy 2494 ATAGGGGTAAGTGTGAGAACTCTCTGATTTGTAATGACATTTCTTATGATGACTTTA 2553
Db 5666 GTAGGGATATAAATTTGAGAAATTTCTCCGATA-----AGTCTAGATGATGATCTCCATA 5719
Oy 2554 TCCATATGCTGCTGATCTTATCCGAATGATCCCAATGCACTATGACAGCTTGTATCAGC 2613
Db 5720 GCTTACGTTCCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5779
Oy 2614 GGACCTCTTGGGAAAGTATATGCAATATCTTACAGACAGAGGCTTTCAGAGTGCATGCA 2673
Db 5780 GGGATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5839
Oy 2674 GGCAGTACACTAGCTTCTCTCTATGTTGAACTGCTGGCGAGTTTGTCTTGAAGTT 2733
Db 5840 GGAATATCATCATGCTTCTTCAACCTTTCAGATTTTCACTGATTTGAAGTGCAGTTG 5899
Oy 2734 CGTGAATCTTCAAGGATTTATATGATGATGATGATGATGATGATGATGATGATGATGAT 2786
Db 5900 CGAGGCTTCTCTGATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5952

```

```

RESULT 10
LOCUS CPN133035 26920 bp DNA BCT 27-FEB-2001
DEFINITION Chlamydia pneumoniae GA1-protein family cluster B, genes omp5,
hb1, hb2, hb3, hb4, hb5, hb6, hb7, omp7, pseudogenes omp8, omp9,
omp15, strain VR1310.

```

```

ACCESSION AJ133035
VERSION AJ133035.1 GI:4455890
KEYWORDS HB1 gene; HB2 gene; hb3 gene; hb4 gene; hb5 gene; hb6 gene; hb7
gene; omp15 gene; omp6 gene; omp7 gene; omp8 gene; omp9 gene; outer
membrane protein 6; outer membrane protein 7; pseudogene.
SOURCE Chlamydia pneumoniae.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 26920)
AUTHORS Daugaard, L., Hjerno, K., Knudsen, K., Madsen, A.S., Christiansen, G.
and Birkebaard, S.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 26920)
AUTHORS Boesen, T.
JOURNAL Direct Submission
TITLE Submitted (21-JAN-1999) Boesen T., Department of Medical
Microbiology and Immunology, University of Aarhus, The Bartholin
Building, DK-8000 Aarhus, DENMARK
FEATURES
location/Qualifiers
1..26920
/organism="Chlamydia pneumoniae"
/strain="VR1310"
/db_xref="taxon:83558"
498..3266
/gene="omp6"
498..575
/gene="omp6"
498..3266
/gene="omp6"
498..3266
/gene="putative"
/codon_start=1
/transl_table=11
/product="Outer membrane protein 6"
/db_xref="GI:4455891"
/db_xref="SPTREMBL:O924H9"
/transl_table="MRSF:CGFPLVSLTLLSFVDSLSATTTSLPEPDSFHDSONA
ERYVNOAGDVSLTGPVSIISVNDNSLNKCFNVSGVSTPGNHHGLFENNISST
TRKGVATCODPQATPARGSEFTLSPTSGSDIKKQCLYSKNMLNNVYVFEON
OSKTKGALISGANTTYGNDSYFTONATFEGATIHSSPDIQAVQATIRPQNTA
KNSGGALYSDDIDIDONAVLFEKEALTTAIGKGAACCLPTSSSTPPLVTS
DNKQLEVERNHSIMGGAITARKLISSSGGPTLEINNI SYANSONLOGAIAIDPGEI
SLSAEKGITFPGNRTSLPFLNGIHLONKFLKQANCSIEFYDPISEADSGSTO
LNINGDKNEYGTITLSEKSLANDPREFKTIPONVMLSGYLVIKGAETVAK
FTOSPGSHVLDIGTKLISKEDIAITGLAIDIDSSSTAAVYKATNKOISVD
STELISPTGNAYEDLKRNSQTPPLSLFEGAGSSTVTRGDELPSPHGROSKML
AMTGTGNKVEEFWDKINYPREKESNLVPLIMGNAVRSIMOVQETHASSLOTD
RGLMIDIGNFEFVHASEDNI RYRHSNGVPLVSVNMEITPKHYSNAFSQLFESKDY
AVSNNEYRMVLGSYLYOYTTSLCNIFRYASRNPNVYGLISRFLOPMLMIFPLCLAY
GHANDMKTDYANP PAVKNSRNRCNAIECGSMPLLVFENGRLFOGAIPEFKLOLVY
AYOGDPEKTTADGRPSNGSLTSISVPLIGTRPERKIALISODVIXDFESSYIPDIPEKDP
SCENALVTSIGDSWLVPAHVSRAHAFVSGTGRYHENDYTELRGISTECPHARNYNI
NCSSKRFPE"
3976..6870
/gene="hb1"
3976..6870
/gene="hb1"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB1 protein"
/protein_id="CA837076.1"
/db_xref="GI:4455892"
/db_xref="SPTREMBL:O923C8"
/transl_table="MRLRLIAVLIALSIIILAGGVLLTVAIPGLSSVISPPAGG
ACALGCMALGIDVLKREVPVIVLAVSTTGTGSPRSGISISGADSTIRISPTYL
LDEGHQSMRKRLIAVLIVFSIIILASGVLLTVAIPGLSSVISPPAGMCAALGC
VMAIGIDVLKREVPVIVLAVSTTGTGSPRSGISISGADSTIRISPTYLDEGHP
OSMRKRLIAVLIVFSIIILASGVLLTVAIPGLSSVISPPAGMCAALGALG
IDVLKREVPVIVPAPIPEEYVIDIDESIRLODAEALAKRLPEMSAFECYTIY
VESHLNEMKSLPYDGHLEKETHQIRVSVSSLKAMPPEFLDIRITEEKEFPFLSKR
KRLIDATLIVERRKILTELEERNLKRAPFVLYODSIFKRTIDNFEKLAWKFMILSKS
ICRFTIIFENHGVAKSLHKRNAVLLEKVIYRSLSQSYRDIGMSAKMRLIHNPFF

```

```

SLEDNKKTKKEHAEMLESISYRKVFLALSDENVDPSPDKMDLSGIPCRDALSE
ISDEOWOKRAHDKHOESLYTOARDRLTQDSSKNOKELEKAQEIYISSWEYKKEPI
ERHETRAIOKLYPNILEREETGOETVTPVOGTASSDLTDILIGLEEVSRKEBN
ONESCVCYKRSHEVMSMEVKOEYGPKEPODMSLEFPEHEEELIYEVLOKDS
KHLSYKRVKNNKEVOYAKRFLKVLBESDLEGIILAQTESASLLTQEBELPLIARALE
KAVFGSLCCALASKAKAPVEEDPRFODSDTQLRALTLRLOEKASLEEBIKRSNLE
NDIAEBERRLKESKOTFERAGLVLREIAYESTYDLSLTNTWEGTPESKRYVSMYL
NYYNEERKRAKRLVETQRYRDFKMLFEMQFNEALIQEELSIOAPSE"
7150..8058
/gene="hb2"
7150..8058
/gene="hb2"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB2 protein"
/protein_id="CAB37077.1"
/db_xref="GI:445893"
/db_xref="SPTREMBL:092503"
/translacion="MOVFLSPOLPPPPRSILGVSCSLRSLSITLLVGLVLLTIGIP
GLFAGISFGALGFSALGVLSGLFLVREVPVPRSEIIPRGVSVPSEPALE
KAKPEETKILDRLPKELDQIDLTIOEVFACLERLKDPKYEEDRGLITTEAKERLAVD
VVKDKMSEPLDQRLVNEBAYIVHERQDPLENIAYETPSSQELRDYTCAGVGYLPS
GDARADLRKRSVEVMDREMRVTWKSWEASVMDHSYGARLEFKKAVGLEBSYKI
LFRSYRDAFYECERKAIQRDGRKWL"
8185..9492
/gene="hb3"
8185..9492
/gene="hb3"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB3 protein"
/protein_id="CAB37078.1"
/db_xref="GI:445894"
/db_xref="SPTREMBL:092300"
/translacion="MWLDRYADKFLIREKEKMERHLEFATMVKASGHAIVAKAKAA
FEKERSNENORRYKDVCKWLSKGLAEFRNOSRRARERLRELQTLPEVSVEERVLER
OFTKYNLENTLYADIEKKYHCVROEHHYKAEYRENGEKVLSAEVSECIQ
RLDCLETWETSKILKAESEVFEKMPATEKLGKVLSDVNRLEIICEDAEEMIFRIE
EITMTLAMEPLILPKNTPEKASLOYNCKSEMLAKYEPCKESPITYRSQWELERLN
ODIATVINCQERLQGSDESCKVTRCDRLREOMKHFEVQGLNFINEELWGAELT
TQARLDVATVPYMEFYLOYHNIRKREVSQWAKTEREIRIQAFQGVAKEDLLAED
TILKEEDYMLLRDWMILRDERKNRRLCNKTAQAQRYKGF"
9805..10698
/gene="hb4"
9805..10698
/gene="hb4"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB4 protein"
/protein_id="CAB37079.1"
/db_xref="GI:445895"
/db_xref="SPTREMBL:0923B1"
/translacion="MLLISGALFTLIGIPGLSAISFGISGIGLSALGVIMTISGLIC
LIVKREIPIYRPEIEGVSLAPSEBPAQAQTLAQLPELDQIDTDOIEVFACIR
KLKDSKYESRSFLNDAKKELRPDVEVDLSEIFELROIYVAGEMDLNGLINGSGL
MMTAESLSLDLFEVSKRILGYLPSGVRGELIKRSKAEIYARLMSLICEIHKVAVADR
NSYNAEKAPAKRLGLAESVYRSLOSTRDKFLSEBRKATIPNGHITWLRDQKSGC
AEKRLQMPRVNGNLKQSPFG"
10776..12119
/gene="hb5"
10776..12119
/gene="hb5"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB5 protein"
/protein_id="CAB37080.1"
/db_xref="GI:445896"
/db_xref="SPTREMBL:0929G3"
/translacion="MDEITFHELYEKTTLKRLHRKCALAKTTFEKKRSKNLQAVEE

```

```

ANARLKYVWDWYDQEFQKAGERLEKLAHALYPEVSVSIRENKIQETRSNLEKAYALE
ENYRCVREQEDWYKBEKREAREFERRGNKILSPBELSSLEQDFDGLKNFSKIMEL
EGLHILQKQATLEVENKILSDAESRLYEIVFEEDWEMPCRIEIEIKTLRMAELPLPT
KKAPEKRSQVNSYKALKEVRKYCKESLAVYVSKERLVSILDRATRYAEQCKRRQ
DSGLSESVRCRQRLREIRIOEPFETOGIDLVKELKLYSSSLRNTREDCVSGYKRAP
GKRFVQYDIELIRVRQSWMTWSEKLRGVQACNMKLAGSEEDKVLKEEYMLY
REERKNKRELVCTIVATQRIQFOPSDIVESSSEKYSIMKARLFFNRHDS"
12262..12987
/gene="hb6"
12262..12987
/gene="hb6"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB6 protein"
/protein_id="CAB37081.1"
/db_xref="GI:445897"
/db_xref="SPTREMBL:092382"
/translacion="MTAAPAILHVSPPPEETKFVIPKDSKRALGILLVGLIVV
CGAIVLGVISGLSALIVCGIGISTISLGVLFVLGILILRRRELTLEQIEAKQIAE
TFPADEKLEMYIOSTEKSLERIEGSRVQGLNARATORIIDLESLSITSEFRL
ROLFDEKTEILSGERLPERIAANLFKQGRDYLNLGNLADIRAYGPNNYKAMVIE
KAKAVYHEPIVLTIMARELEFF"
13014..14585
/gene="hb7"
13014..14585
/gene="hb7"
/note="hypothetical"
/codon_start=1
/transl_table=11
/product="HB7 protein"
/protein_id="CAB37082.1"
/db_xref="GI:445898"
/db_xref="SPTREMBL:0956P3"
/translacion="MYQENLRILRLRYLNSVOKSYADRLSEYKTKVMDPLIPWEE
DKERCAEAKALEQOKLILIDYGKSTFWLMEANDILNDPWSGLNTVPRKYFOED
DSERWMHVKVLIQLEQNYEKLLESSKESTEAKKLESLFLPLSIEBELVQCLEDDIYKKE
VETRVLDARAYGVTDPPKODTEPAKKKVELASLEFLPLSIEBELVQCLEDDIYKKE
ODVDKDLARQLEQODIEAKREBAADRLNRLKSKTMDLRAMHLENAEDSLTM
WTSQILNKQDKARKLILKEDITSVLPIDEIEFTCLSLBELPLITRELLTKSVLKRI
CSFTLAKMYSVFENNITYOYEVYQVONLGFKLQGISORQKQKDDPANLEQVALOK
RLBELQNEFTIGFNFMEKDEKAAKADLYIRSTAEQKMNPDVPCHELFRYHEVVKP
LLEIMVNCADSYNDARKKLSLRDLDEKLELQKKEEFYOKKQORHADRSHRTYQK
LREIYALELAKKI"
14954..17479
/gene="omp7"
14954..17479
/gene="omp7"
/note="putative"
/codon_start=1

```

Query Match 23.64; Score 663; DB 2; Length 26920;  
Best Local Similarity 55.68; Pred. No. 2,3e-148;  
Matches 1365; Conservative 0; Mismatches 1165; Indels 83; Gaps 12;

```

QY 1 ATGAATCGCAATTTTCCTGGTAGTCCTCTCGACATTTGGCATGTTTACTAGTTGT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20786 ATGAGATCGTCTTTTCTGTTATATCTCATCTGACGCTTTCCTCTATATG 20845

QY 61 TCCACTGTTTTCGTCGACATGCTGAAAATRAGGCCCTCTGATACCTTTGACGCAAGT 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 20846 AGTGTTCCTCGAGATGCTGCGATCTCACAATTAGGAGTGTGACACTTAAAGTGAT 20905

QY 121 ACTACACAGGACCACTTACTCCTTAAATAAGCACTAC-----TGGAAATGACTAT 171
    || | || | | | | | | | | | | | | | | | | | | | | | | | | |
Db 20906 ACAAGCACACAGATTTACTCTTAAGCGGCACTTCTGATGCTAATGGCAGCACTAT 20965

QY 172 ACTCTGACAGAGATATTAAGCTGCAAAACCTTGGGATTCGCGACGCTTAAACGAAGGT 231
    || | || | | | | | | | | | | | | | | | | | | | | | | | | |
Db 20966 ATTCTGATGGGATGCTCGATTAAGCCAGGAGGAAACAAAGACACTTAACCAAGAT 21025

QY 232 TGTGTTTTCGACACTACGGAATCTTTAAGCTTTGGCGGTAAAGGGTACTACTTTCTTT 291
    || || || || || || || | | | | | | | | | | | | | | | | | |

```



Dh 21026 TGTCTTCTACACTGACGAAATCTTACCTTTAGGGAACGATTTCTTCATTTT 21085  
Qy 292 TTAATATTAAAGTGTAGTGTGAAGGCGCAGACTTCTGTACAACTGATAAAAATCTG 351  
Dh 21086 GACAAATATATTTCTCTACT- - -GTTCAGAGTGTGTGTAGCAATACAGCACTTCT 21142  
Qy 352 TCGCTACAGAGATTTTCGAGTCTTACTTTCTTACGGGCCCATCATCGGTATGCACACC 411  
Dh 21143 GGGATATTCAAAATCTCAGAGATTTTCAACCTCTCGAGATGCTTGGACGCTCAGGA- - - 21197  
Qy 412 CCCCTCAGAAAAAGTGCAGTAAATGTGAGGGGATCTTACATTTGATTAACAAATGGAAT 471  
Dh 21198 -CCACAGTAAAGAGGACCTTAAATTAACCGATGTCTGTGTGTAGAGATAGGGGAT 21256  
Qy 472 ATTTATTTAAACAGATTACTGTGAGGAATGGGAGGACATTTCCACAGAAATCT 531  
Dh 21257 CTGTATCTTAATGAATAATGCTCTAGTGAATAATGGGGAGCCATACAAAGACTTTG 21316  
Qy 532 TCTTTGAAAAACAGCAGCGGATTCGATTTCTTTGAAAGGAATTAATCGAGCGCAACAGG 591  
Dh 21317 TCTTTGACTGGGAGTACCGGGTTTGTACCGTTCCCTGGCAATAGCTGTC- - - - -G 21367  
Qy 592 AAAAAAGTGGGCTATTGTGTCTACTGTACTGTAGATTAATTAATTAATAGCGCTCT 651  
Dh 21368 CAAACAGGGGAGGATCTATGTCTTGTGACTCTGTGATTTCTGAANAATGACGAATC 21427  
Qy 652 ACCCTTCTCGAACAAATATTGCTGAAGCTGACGAGTGAGGTATTAATAGCAAGAAC 711  
Dh 21428 TTGAGCTTCGAAACACAGTCCGACAACTACAGAGGCGGATCTGCTGAAGGGAAC 21487  
Qy 712 TGTCAATTTACAGGAATACGCTCTGTGTATTTCTGAAATAGTGTACAGCACCGCA 771  
Dh 21488 CTGTGATCTCCAAATACCAAAATATCTTTTTCGATGGCTCAAAAGCAACTACAAATGCG 21547  
Qy 772 GGA- - - - -AATGAGAGCTCTTTGTGAGATGCCGATTTACCAATCTGGG 819  
Dh 21548 GGAGCTATTGATGTGTAACAAAGGCGGACACCGACCTATCTGACTTTTCAGGA 21607  
Qy 820 AATCAGAGTGTACTTCTCAGGAACCAACGCTGATGATATGGCGGACCATTTATGCT 879  
Dh 21608 AATGAGACCTGCAATTTCTGAATTAACACAGCAAGAAATATGTGAGGTCGCAATTTATAC 21667  
Qy 880 AAGAGCTTACACTGCTTCCGGGGGGGGGGTATCTCTTTTCAACAAATATAGTC 939  
Dh 21668 AAAAAATGTGTATCTCTCAGAGAGAGAGAGTGTATTTTCTAACACAAAGCTGCG 21727  
Qy 940 CAAGGTACCACTGAGTATGTGTGAGCCATTTCTATCTGAGCACTGAGAGTGTAGT 999  
Dh 21728 AATGCTACTGC- - - - -TAAAGAGGGGCAATGTGCAATTTAGATTTGAGAGATTAGC 21781  
Qy 1000 CTTTCAGAGAGAGGAGGACATTAACCTTCATAGGAATGCCATTTGTGCAAT- - -ACA 1056  
Dh 21782 ATTTCTGAGATCTCGCAATATCATTTTGGAGGCAATTAATCTAGACACATCAAGAGT 21841  
Qy 1057 CCACAACTACAAAAAATTTCTATTGACATAGATCTACTGCAAAAGATCACAAATTTA 1116  
Dh 21842 CCTCGAGTGTGACACAGAAATGCTATAGATCTTGCATCGAATGCAAAATTTTAAATCTC 21901  
Qy 1117 CGTCAATATCTGGGCTAGCATCTTTTCTAGATCCGATTTACTGCTAATAGCGCTGCG 1176  
Dh 21902 CGAGCGACTCGGGGAAATTAAGTATTTTCTATGATC- - - - -CTATACGAGCTCA 21952  
Qy 1177 GATTCACAGATCTTAATTAATCTCAATAGGCTGATGAGTATATAGTATATAGT 1236  
Dh 21953 GGAGCTACTGATAGCTCTCTTTGAATTAAGCTACGAGGATGTGGAAATACCTATGAA 22012  
Qy 1237 GGGTCGATTTTTTTCTGTGTGAAGCTCTCTGAGATGAGCAAAAGTTGACAGAAC 1296  
Dh 22013 GGGTACATCGTTTCTGTGGAGAAACTCTCAGAGAGAACTTAAGAAACCTGACAAAT 22072  
Qy 1297 CTCACCTTACAGCTGAAGACGCTGTACTTAACCTGACGAGAAATTTAGTACTTAAAGCT 1356  
Dh 22073 CTGAAGCTTACATTTACACAGGCTGTAGAGCTTCTGAGGTCGCTTATGTAATGAAGAT 22132

Qy 1357 GGTGCACTCTGATACGAAGGCTTTTACTCAGACCGGGGTCTCTGTTATTATGAT 1416  
Dh 22133 GGAGTACCTGATGTGCAAAATCTATTAACGAGCTGAGGGATGCAAGTGTATGAT 22192  
Qy 1417 GCGGCAACAGCTTAAAGCAAGTACAGAGAGTCACTTTAAACAGTCTTCCATTCCT 1476  
Dh 22193 GGAGGACTACTTTTGAAGCAAGCGCTGAGGGGGTCACTCTCAATAGGCGCTAGCTTAAT 22252  
Qy 1477 GTAGACTTTAGGCGAGGGTAAAGAAATGTATGCTGCTTCTGCGACAGTAAAT 1536  
Dh 22253 ATAGATCTTAAAGGAGCAAAATTAACCTATCAATTAAGGCGAGCGCAAGTAAAGAT 22312  
Qy 1537 GTAGCCCTTAGGTCGCAATCTTCTTGTGATTAACAGGGAATGCTTAAGAAATAC 1596  
Dh 22313 GTTGCCTTATCAGGGCTCTATATGCTTTAGATGCTCAGGGGAACCTATTAAGAGTAT 22372  
Qy 1597 GACTTAGGAAAACTCAAGACTTTTTCATTTGTGACAGCTCTGCTGTGCTGACTGCAACA 1656  
Dh 22373 AATCAGATCAACAGCAGTCTTCTCTTAATAGAGCTTTCTGCAACAGAACTGATGAT 22432  
Qy 1657 ACTACAGATGTTCCAGCGGTTCTACAGTACCAACTCTTACGACTATAGGTATCAAGT 1716  
Dh 22433 ACTACAGATATCCCGATACCCCAATCTAATATCACTACGAATCACTATAGGTATCAAGG 22492  
Qy 1717 ACTTGGGGAATGACTGGGTGATGATATACCGCAAGCACTCCAAAGACTATAGACAGCGACA 1776  
Dh 22493 AACTGGAATTAATTTGTTGGTGC- - - - -ACGATGCACCTGCAAAAACAAATATCTAC 22547  
Qy 1777 TTAGCTTGGACCAATACAGGCTTACCTCCGAATCCGATCCGATCAAGACCTTTAGTTCCT 1836  
Dh 22548 TTAATCTGATTAACAGAGATACAGCGCAATCCAGAACTCAGAGCTCAGGAGCTTTGGTTCCT 22607  
Qy 1837 AATACCTTGGGATCTTTTACAGACATCCAGAGATTCAGAGTGTATAGAGAAAT 1896  
Dh 22608 AATAGCCTGTGGGTTCTTTGTGATGTCCGCTCAATTCAGAGCTCATGAGCGGAGC 22667  
Qy 1897 GCTTGACTCTTGTGATGATCGAGGCTTGGGCTGGGGAGTCCCAATTTCTTAAT 1956  
Dh 22668 ACAAGTCTGTTATCTGCTCAACAAATTTGTGGGATAGGAATTCGAGCTTTTTCAT 22727  
Qy 1957 AAGATTAAGAAAGGGGAAAAACGCAATACCGTATAATGTGTGATGATATCGA 2016  
Dh 22728 GAAGATCAGAAAGAAACCAACAGTATGATCTTACGCGGGTATATGCAATAGGA 22787  
Qy 2017 GGTGCAAGGCAACTTGTCTGAAAACCTTAATAGCTTGTGCTTTTGCAACTCTTGGT 2076  
Dh 22788 GGAGGATTTCTCAGCGCTTCTGAAAATTTCTTAATTTTGTCTTTTGGCACCTTTTGGC 22847  
Qy 2077 AGCATTAAGATTTCTATAGTGGCTTAAATATCATGATFACCTATGACGAGGCTTCTAT 2136  
Dh 22848 TAGACAAAGGACATCTTGTGTGTAAGAACCTACCCATGATATGACGGGCAATGAGT 22907  
Qy 2137 ATCCAACATTAAGATAGTGTAGTGGGTTCAATAGTGTCTCTTAAGATTAACCTCTGGC 2196  
Dh 22908 TAGCGAACCTCGGAGAGTCA- - - - -AGACCCGCTGAAGATTTTGTGACAGA 22955  
Qy 2197 TCTTGAATCATTAACCCCTGTTTGAAGGGCAGCTGCTTATAGCCAGCTAGTAAAT 2256  
Dh 22956 AATTCGACTCCCACTCTTGTCTTCAATCTCGGTTTGTCTTATGCGCATACGACAAAT 23015  
Qy 2257 GATCTAGAGCAAAAGTATACCGGTATCCGATCCGAGGAAAGTCTTGGGGGAATTAATGCT 2316  
Dh 23016 AACAATGACCAAAAGTACACTGGCTATCTCTGTATAGGGAACCTGGGGAATGATGCC 23075  
Qy 2317 TTTAACAATGATGTGGAGCTTCTTCATCTTATCT- - -GAATACCTGCAATGTGTTT 2373  
Dh 23076 TTGCTATAGAAATGTGGAGGATATCCCGGTAGTGGCTTACAGACGTCGGTCTTGCGTG 23135  
Qy 2374 GATACCTATGCTCATATACATTAACGTAAATGTGACCTATATAGCTCAGAGACGTTCTCG 2433  
Dh 23136 GATACCCACAGCCATTTCTAAACCTTAGAGATGTATGACATCAAGAAATGACTTTAAG 23195



OY	2434	GAGAAAGGTACAGAGAGAAAGTCTTTTGTATGACAGCAAGCTCTTCAATTATCTTGGCT	2493
Db	23196	GAAACGGCACAGAGAGCCGCTTCTTCCAAAGTCAGAACCTCTTCACTACAGCGGTTCT	23255
OY	2494	ATAGGGGTGAAGTTTGAGAACTCTCGATCTGATCTTAATGACCTTTTCTTAATGATCTGACTTA	2553
Db	23256	GTAGGATTAATAATTGGAGAAATCTCCGATA-----AGTCTAGATGATCTCTCCATA	23309
OY	2554	TCCATGTTCTCGATCTTATCCGCAATGATCCCAATGACATCCAGCTACAGACTCTTAATCAGC	2613
Db	23310	GCTTACGTTCCCGATGTCATTCGTAATGATCCAGCTGCACGCAACCTCTTAAGGTTTCT	23369
OY	2614	GGAGCCTCTTGGGAAACATTATGCCAATTAACCTATAGCACAGAGCCCTTGCAGAGCGCTGCA	2673
Db	23370	GGGGAFTCTTGTCGACATGTCGACAAAGCTTGTCTGACAAACCTCTCTTCTGACGTGCT	23429
OY	2674	GGCACTCACTACGCCCTTCTCTCCATGTTTGAAAGTGTCTGGCCAGATTGTCTTTGAAGTT	2733
Db	23430	GGAAATCACTCATGCTTCTTGCCTTCAAACTTGAAGTTTCAAGTCTTGAAGTCAGATTG	23489
OY	2734	CGTGATCTCTCAGCAGATTATATGATGATCTTGGGGGTGAAGTTCCCAATCTCA	2786
Db	23490	CGAGGTTCTTCTCGTACGTATGCTATGATCTTGGAGGAGAAATTCGGATTTTA	23542
RESULT	11		
LOCUS	AE002193/c		
DEFINITION	AE002193 10026 bp DNA	BCT	30-MAY-2000
ACCESSION	AE002193		
VERSION	AE002193.2		
KEYWORDS	GI:8163406		
SOURCE			
ORGANISM	Chlamydia pneumoniae AR39.		
REFERENCE	Chlamydia pneumoniae AR39.		
AUTHORS	Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.		
	1 (bases 1 to 10026)		
	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,		
	White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,		
	Berry,K., Bess,S., Linher,K., Khouri,H., Craven,B.,		
	Bowman,C., Dodson,R., Gwin,M., Nelson,W., Deboy,R., Kolonay,J.,		
	McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.		
	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia		
	pneumoniae AR39		
JOURNAL	Nucleic Acids Res.	28 (6),	1397-1406 (2000)
MEDLINE	20150255		
PUBMED	10684935		
REFERENCE	2 (bases 1 to 10026)		
AUTHORS	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,		
	White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,		
	Berry,K., Bess,S., Linher,K., Khouri,H., Craven,B.,		
	Bowman,C., Dodson,R., Gwin,M., Nelson,W., Deboy,R., Kolonay,J.,		
	McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.		
	Direct Submission		
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712		
JOURNAL	Medical Center Dr. Rockville, MD 20850, USA		
COMMENT	On Jun 1, 2000 this sequence version replaced gi:7189234.		
FEATURES			
source	Location/Qualifiers		
	1..10026		
	/organism="Chlamydia pneumoniae AR39"		
	/strain="AR39"		
	/db_xref="taxon:115711"		
gene	/note="synonym: Chlamydia pneumoniae AR39"		
	complement(105..2897)		
	/gene="CP0307"		
CDS	complement(105..2897)		
	/gene="CP0307"		
	/note="similar to GP:4376729; identified by sequence		
	similarity; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic membrane protein G family"		
	/protein_id="AAF38166.1"		
	/db_xref="GI:7189237"		
	/transl_table=11		
	/product="polymorphic		

/db\_xref="GI:7189235"  
 /translation="MKRIPLKRLILSTVTPTLISLITAYGADASLSLPTDSFGAGSGST  
 ETPKSTADANGNTYLSGNYIINDAGKRTALTGCFFETTGDLPTFGKYSFSPNTYD  
 FPKSGTAADKALPTFGSLNLSIFAAGPTVAKSLTSSGALNTLNDKPTLTD  
 SOVNSAGNAANNNGAIYATKTLISGNTSITPTSPNKAAGGAIYSKAASISGNTGDL  
 VNNKSGTEFGGALGFEEASSITGONSLSFEGEAGGALDAGGAIYCEKTEIDPTLTI  
 SPMKSTPTFENGSGVGGALCAHGDLSAAGPTLFENNRCPNTFAAGKAIALDSSCS  
 LLSANDOGITFLGTLTSTSPATSTRAIYAGSAGATKTNLRAAGOSIYDPTIASNL  
 TGLSDVGLITINDPSNPLDYSCTIVEGKLSLDEAKAANPTLSIKOPALASGLN  
 AKGNVEIDNVEFTOTBESSTLMOPGTLKADPTAISTRTKYVNLSEKGSISFTY  
 AAKNTITLTSPLVFODSGNFTBSHTINQAFPTOPLVFTATYAADSIYDALTSYV  
 QPEPEHYVGQGWEMATWADTSTAKSGIWTWDTGNPNPERASAVDPDLSASTDTP  
 TLQOITSCGANSIYOORGLMAGSGTANFPHKRSQGNQAFKMSFGSYIYGSAGDESPNS  
 IYVSVAFCQSGEKDPLFIVENTSHYLAISLQHAEPGLJPMFSFGITPTMLDIDPL  
 LINDYSTYTKNDMDRTYSPEAGSGDGLVNCSSIEGLGSLAIYLPREAFPGCYEV  
 FLKFOAVYSROONREKESGAEAPAPDDGVLNVCSTPYGIRLEKISDEKDNNEISLAIYI  
 GYVTRKRPNSRTSLMVSAGSWTSLCKNLARQAFSLASGSHLTSPLHYELSGEAAIYELR  
 GSAHIYNDGGLRYSF"  
 complement(2941..5751)  
 /gene="CP0308"  
 complement(2941..5751)  
 /note="similar to GP:4376720; identified by sequence  
 similarity; putative"  
 /codon\_table=1  
 /transl\_start=1  
 /product="polymorphic membrane protein g family"  
 /protein\_id="AAE38165.1"  
 /db\_xref="GI:7189236"

gene  
 CDS

/db\_xref="GI:7189235"  
 /translation="MKRIPLKRLILSTVTPTLISLITAYGADASLSLPTDSFGAGSGST  
 ETPKSTADANGNTYLSGNYIINDAGKRTALTGCFFETTGDLPTFGKYSFSPNTYD  
 FPKSGTAADKALPTFGSLNLSIFAAGPTVAKSLTSSGALNTLNDKPTLTD  
 SOVNSAGNAANNNGAIYATKTLISGNTSITPTSPNKAAGGAIYSKAASISGNTGDL  
 VNNKSGTEFGGALGFEEASSITGONSLSFEGEAGGALDAGGAIYCEKTEIDPTLTI  
 SPMKSTPTFENGSGVGGALCAHGDLSAAGPTLFENNRCPNTFAAGKAIALDSSCS  
 LLSANDOGITFLGTLTSTSPATSTRAIYAGSAGATKTNLRAAGOSIYDPTIASNL  
 TGLSDVGLITINDPSNPLDYSCTIVEGKLSLDEAKAANPTLSIKOPALASGLN  
 AKGNVEIDNVEFTOTBESSTLMOPGTLKADPTAISTRTKYVNLSEKGSISFTY  
 AAKNTITLTSPLVFODSGNFTBSHTINQAFPTOPLVFTATYAADSIYDALTSYV  
 QPEPEHYVGQGWEMATWADTSTAKSGIWTWDTGNPNPERASAVDPDLSASTDTP  
 TLQOITSCGANSIYOORGLMAGSGTANFPHKRSQGNQAFKMSFGSYIYGSAGDESPNS  
 IYVSVAFCQSGEKDPLFIVENTSHYLAISLQHAEPGLJPMFSFGITPTMLDIDPL  
 LINDYSTYTKNDMDRTYSPEAGSGDGLVNCSSIEGLGSLAIYLPREAFPGCYEV  
 FLKFOAVYSROONREKESGAEAPAPDDGVLNVCSTPYGIRLEKISDEKDNNEISLAIYI  
 GYVTRKRPNSRTSLMVSAGSWTSLCKNLARQAFSLASGSHLTSPLHYELSGEAAIYELR  
 GSAHIYNDGGLRYSF"  
 complement(2941..5751)  
 /gene="CP0308"  
 complement(2941..5751)  
 /note="similar to GP:4376720; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=1  
 /product="polymorphic membrane protein g family"  
 /protein\_id="AAE38166.1"  
 /db\_xref="GI:7189237"

gene  
 CDS

/db\_xref="GI:7189235"  
 /translation="MKRISLPMWLITSSALVPSLHPLMAANTDLSDDYENGSSGSAAF  
 TAKETSDASGTYTLTLDSPSLITNVSALTPARKSGPTNGTGLSPVGDHSLVLTQIAL  
 THDGAANNINTNALSFGESLLIDSAPATSCGGALCYTNEGTEATGAGIYETEDPTFLGSGYVAT  
 LOKNTSEKDGANVSAYSIDLAKTTTALLDONTSTKNGALCSTANTYVONSSTYVE  
 SNTTGGGGGTYSEKERTSIDCANTGVPTFSNARKGGAASNDNALINGNOVLV  
 EKKTTGSAOANNEPGCGGATCCTIATPTDKTGTGLAISONSOMSTSTNTTANGAIYAA  
 TKTCTLDGNTLTFDONTIYVAGCGGAIYETEDLSKSGSYVTSTNARKGALVSK  
 GNSSLTGNTNLTFSGNKATGPSNASSANOEGCGGALISFLESASVTKGLMIENENVA  
 SLISGTAATVGGGAIYATKALGHNTLTFDONTAETAGIYETEDPTFLGSGYVAT  
 FSTNATKAGALHTKGTNSFTPKNALVSGSAAATATTDDGCGGALCINSESD  
 IATKTLTLENSELSFINNTAKRSGGGLYAKVYISGESINPGNTANGSAGAIYSEK  
 NLSITANPGPTNNSGKGAIIYASOGLSLAIDGTLPGSCNATBCTSPNSHT  
 LGAGNPKTKLAAPGHTIYTYDPTITMEPAGSGTIBEYINPYVKALVPPQKNGPT  
 ASVPAVPAAPNPNGTIVFSGSKLPDSASIPANTYTLNOKITNLGAVNVLEKGT  
 LOVFSYQDPSYTFMDAGTGLLETTNTNDGSIDLKRLSNTVDNALDGRKRTIYAVNS  
 TSGKLTIGDGLTFHNNEGSPYDNGPLKANMLPFLDLSYSTSGVNLDDFPNIFSSMAV







Dh 12646 CTCCTAGCCAAAAGTCTCCATGAGCTAATAACAATGGCGGAGCAGTACACACAAA 12705  
Qy 526 AATCTTCTTGAAGAAACAGCAGCGGATCGATTTCTTTGAAGGAATAAATCGCGCA 585  
Dh 12706 ACCTTCTTATTTCTGGGAATACCTCTTCTATACCTTCACTAGTAATAGGCAAAA--- 12762  
Qy 586 ACAGGAAAAAGGTGGGCTATTGTGCTACTGTATGTATATTACAAATTAATACG 645  
Dh 12763 -----AAATGAGTGGAGCATATAGCTCGCGCTGCAGATATTTTCAGAAACAC 12816  
Qy 646 GCTCTACCTCTCTCTGCAACAATTTGCTGTAACCTGAGGTGAGCTATTAATAGACA 705  
Dh 12817 GCGCAGGTAGTCTTTATGAAATTAAGAGAACTGGGGTGGGCTTGCTTGAA 12876  
Qy 706 GGAATCTGATACATTAAGAGGAATACGCTCTGTATTTTGAAGAAATAGTGACACG 765  
Dh 12877 GCCAGCTCTCTGATTAATCAAAATAGCTCCCTTTCTCTGGAACACATGCAACAGAT 12936  
Qy 766 ACCGAGAAATGAGAGAGCTCTTCTGAGA-----TGCCGATGTTACC 810  
Dh 12937 GCTCAGAGCAAGGCGGGCCATTTATTTGAAAAAACAGAGAGACTCTACTTACT 12996  
Qy 811 ATATCTGGATATCAGATGATTAATCTTCTCAGGAAACCAAGCTGATAGTAATGGGAGCC 870  
Dh 12997 ATCTCTGAAATTAAGTCTGACCTTCCGCGAAGACTCTTCACTCAAGGCGAGACA 13056  
Qy 871 ATTTATCTAAGAACTTACACTGGCTTCCGGGGGGGGGATCTCTCTTTCTTAC 930  
Dh 13057 ATCTGTCCTCATGCTAGATCTTCCGCTCTGCCCC-----TACCCTATTTTCAAT 13110  
Qy 931 AATATCTCAAGGTACCACTGACAGTAATGTTGAGGCAATTTCTATATCTGCACTGGA 990  
Dh 13111 AATGATGAGCGGGAACAGCTGAGGCAAGGCGCCTATGCAATGCGCATCTGGA 13170  
Qy 991 GAGTGTGCTCTTGAAGAGAGGAGACATTAATCTTCAATGGGAATGCCATTTGGA 1050  
Dh 13171 TCTTAACTCTCTCTGCAATCAAGAGACATCACTCTTCTGCAACACTCTTAACCTCA 13230  
Qy 1051 ACTACACCACAACTACAAAAAAGAAATCTATGACATAGATCTACTGCAAGATCAG 1110  
Dh 13231 ACCCTCCGCGCATGACAGACAGGATGCTATCTACCTGGAGTGTGCAAAAAATATCG 13290  
Qy 1111 AATTAGTGTCAATATCTGGGATAGATCTTTTCTACGATCCGATTAATGCTTAATACG 1170  
Dh 13291 AACTTAAAGGAGCCCAAGGCAATCTATCTATTTATGATGATGCAATGCTTAACACC 13350  
Qy 1171 GCTGCGATCTTACAGATCTTAAATCTCAATAGGCTGATGAGTAATAGTACAGAT 1230  
Dh 13351 ACAGAG-----CTTACAGCTTCTGACATCAACCAACCGGATACCACTGCTTTAGAT 13407  
Qy 1231 TATAGTGGGTCGATTTTCTTGTGAAAAAGCTCTGTGAAGATGAAGCAAAAGTTGCA 1290  
Dh 13408 TATTCAGGAACGATATTTATTTCTGGGAAAAAGCTCTGCAAGTGAAGCAAAAGTCT 13467  
Qy 1291 GACAACCTCACTTACAGCTGTAAGCAAGCTGTACTGATCTGAGGAAATTTAGTACT 1350  
Dh 13468 GATTAATCTCAATTAATTAAGCAACCAATGCTGTAGCTCTGGAACCTTGAACCTC 13527  
Qy 1351 AAAAGTGTGCACTCTGATCGAAGAAAGCTTACTCAGACCGGCGCTCTCTGTTATT 1410  
Dh 13528 AAAGGAATATGCTGAGTATGATGTAATGTTTTCACACAGACTGAAGGCTCTACACTCTC 13587  
Qy 1411 ATGATGCGGAGCAACGTTAAAAAGCAAGTACAGAGGAGGCTACTTAAAGGCTTTCC 1470  
Dh 13588 ATGCAACCAAGAACCAAGCTCAAGATCTGAGCTCTCACTTACCAAACTGTCTC 13647  
Qy 1471 ATTCCTCTACACTTATAGGCGAGGTAAGAAAGTGTATGCTGCTCTGCGCAAGT 1530  
Dh 13648 GTTGAATCTTCTGCTTAAAGGAAATTAAGAGTGTGCTCAATGAAAGAGAGGCCAAC 13707  
Qy 1531 AAAAATAGCCCTTAGTGGATCTCTTCTTTGATTAACCAAGGATGCTTATGAA 1590  
Dh 13708 AAAACTATTAACCTTACCTCTCTCTGTTTTCAGAGATAGTACGGCAATTTTATGAA 13767

Qy 1591 AATCAGACTTAAAGAAAACTCAAGACTTTTCAATTTGACAGCTCTCTCTGGTACT 1650  
Dh 13768 AGCCATACGATAAACCAAGCCTTTCACGAGCCTTTGGGATTCACATGCTGCTAGTCT 13827  
Qy 1651 GCAACACTACAGATGTTCCAGCGGTTCTTACAGTACCAACTCTTACGACTATGGGAT 1710  
Dh 13828 GCTAGCATATATTTATTCGATGATGCTCTCTCTCTCAGTACCAACTCCAGAACCTCAT 13887  
Qy 1711 CAAGTACTTGGGAATGACTTGGTTGATGATACCCGAAGCACTCCAAAGACTPAAGCA 1770  
Dh 13888 TACGGGATACAGGA---CATTTGGGAAGCCACTTGGGACAGACATCACTGCAAAATCA 13944  
Qy 1771 GCGACATTAGCTTGGACCAATACAGCTTACCTTCCGATCCTGAGCGTCAAGACCTTTA 1830  
Dh 13945 GGAACATGACTTGGTAACAGCTGCTTAACCAACCTTATCTGAGCGTGAAGCTTCCGTA 14004  
Qy 1831 GTTCTTAATAGCTTTGGGATCTTTTTCAGACATCCAGAGATTCAGAGTGTATGAG 1890  
Dh 14005 GTTCCGATTCATTTATGGGCACTCTTACTGACATTCGCACTTACAGCAGATCATGACA 14064  
Qy 1891 AGAGTCTTGTGACTTTGTTGATGATCGAGGCTTGGGCTGGGAGTCCCAATTTG 1950  
Dh 14065 TCTCAGGAAATATATATCTATCAGCAAGAGACTCTGGGATCAGAACTGCGAATTTG 14124  
Qy 1951 TTGATTAAGATTAAGAAAGGGAAGAAAGCAATACCGTATTAATCTGTGATATGCT 2010  
Dh 14125 TTGCATTAAGATTAATCAGAAATACCAAGCAATTCGACATTAAGCTAGGCTATATTT 14184  
Qy 2011 ATCGAGGTGACGCGCAACTTGTTCGAAACTTAATAGCTTTGCTTTGGCACTC 2070  
Dh 14185 GTTGGAGAAAGTGTGAGATTTTCTGAAATATCTCAGTGTAGGCTTCTGCACTC 14244  
Qy 2071 TTTGATGAGCATTAAGATTTCTTGTGCTTAAATATATCTGATTAATCTGACGAGACC 2130  
Dh 14245 TTGCTTAAGATTAAGACCTGTTTATAGTTGAATACCTCTGATTAATTTAGCGTGG 14304  
Qy 2131 TTTATATCCACATTAACAGAAATGATGAGGTTCAATAGGTTCTCTTATGATTAACCTT 2190  
Dh 14305 CTATACCTGCAACATCAGCACTTCTAGGAGGACTTCCATGCTTATTTGGAAGTATC 14364  
Qy 2191 CTTGGCTCTTGGATCAATTAACCCCTGTTTGAAGGAGCTCGCTTATAGCAGCTC 2250  
Dh 14365 ACCGACATGCTGAAGATATTCCTCTCATTTTGAATCCCGCTTAAGCTTACAGTACT 14424  
Qy 2251 AGTATATCTGAAGACAAAGTATATCTGCTATCTGAGTGAAGGTTCTTGGGGGAAT 2310  
Dh 14425 AAAATGATATGATCTGCTATCTGCTATCTGCTATCTGAGCTCAAGGCTCTGTGACCAAT 14484  
Qy 2311 AATGCTTTAATGATGATGTTGGAGCTTCT---TGTCTTCTTATCTGTAATACCTGAT 2367  
Dh 14485 AACTCTGGGCTCTAGAGCTCTGAGGATCTCTGCTCTATATCTCCCTAAAGAGCACG 14544  
Qy 2368 TGTTTATATACCTTATGCTTCCATATCAATCAAACTGATCACTTATATAGCGAGGACG 2427  
Dh 14545 TTTCTTCCAGGATATTTTCCCTCTTTAAAGTTCCAGGCACTTACACCCGCCCAAAAC 14604  
Qy 2428 TTTCTGGAAGAAAGGTACAGAAAGATCTTTGATGACAGCAACCTTCAATTTATCT 2487  
Dh 14605 TTTTAAAGAGTGGGCTGAGAGCCCGTGTGATGATGATGAGACCTTAAGTATGCTCT 14664  
Qy 2488 TTTGCTATAGGGGTGAAGTTTGAAGATCTCTGATTTGTAATGACTTTTCTTATGATCTG 2547  
Dh 14665 ATCCCTGCTGCGATTCGTTAGAAATAATCTCCCAATATTAATTAATTTTCGATTT 14724  
Qy 2548 ACTTATCTATGTTTCTGATCTTATCCGAATGATCCGAATGCACTCAATGACAGCTGTA 2607  
Dh 14725 TCTTACCTATACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 14784  
Qy 2608 ATCAGCGAGCTCTTGGGAAACTTATGCAATTAATCTACAGCAGAGGCTTGAAGTG 2667  
Dh 14785 GTGAGTGGAGCTCTTGGACTTGGCTATGTAATAAAGCTCCGACAGCAAGCTTCTTATGCA 14844

Oy	2668	CGTGGAGGACATCTGACATGCGCTTCTGCTCATGATGCTGGCCAGCTTTCCTT	2727
Db	14845	AGTGTGGAAGCCATCTGACTCTCTCCCTTCATGTAAGACTCTCTGGGAGACTGCTTAT	14904
Oy	2728	GAACTTCGTGATCTTCACGAGATTATATATGATAGATCTTGGGGGTAACTTCCAAATCTAG	2787
Db	14905	GAGCTTGTGCTGCTGACACATCTACATGTAATGTTGGCTAAGATACATCTCTAG	14964
Oy	2788	GAGCGTCTTCATGCTGCA	2806
Db	14965	TTTCCTACTTTCCTCCCTAA	14983
RESULT	13		
LOCUS	AP002546		
DEFINITION	Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,	BC1	08-DEC-2000
ACCESSION	AP002546	300650 bp	DNA
SECTION	2/4		
AB0033802	AB0033782	AB0033783	AB0033784
AB0033803	AB0033804	AB0033805	AB0033806
AB0033807	AB0033810	AB0033811	AB0033812
AB0033809	AB0036080	AB0036081	AB0036082
AB002546.2	GI:10176692		
KEYWORDS			
SOURCE	Chlamydomophila pneumoniae J138 (strain:J138) DNA.		
ORGANISM	Chlamydomophila pneumoniae J138		
REFERENCE	Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.		
AUTHORS	Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K., Shiba,T., Ishii,K., Hattori,M., Kohara,S. and Nakazawa,T.		
JOURNAL	Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA		
MEDLINE	Nucleic Acids Res. 28 (12), 2311-2314 (2000)		
REFERENCE	20330349		
AUTHORS	2 (Passes 1 to 300650)		
TITLE	Shirai,M.		
JOURNAL	Direct Submission		
COMMENT	Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases. Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan (E-mail:mshirai@pc.yamaguchi-u.ac.jp, Tel:81-836-22-2227, Fax:81-836-22-2415)		
On Sep 15, 2000 this sequence version replaced	gi:6172290		
gi:6172282	gi:6172294	gi:6172286	gi:6172326
gi:6172332	gi:6172334	gi:6172336	gi:6172338
gi:6172388	gi:6172388	gi:6172390	gi:6172392
gi:6174668	gi:6635174	gi:6635176	gi:6635180
gi:8547438	gi:8978640.		
AB0033782-AB0033785,	AB0033800-AB0033815:		
AB0038348-AB0038349:	Submitted (14-Feb-2000)		
AB0036079-AB0036082:	Submitted (18-Dec-2000).		
location/Qualifiers			
1. 300650			
/organism="Chlamydomophila pneumoniae J138"			
/strain="J138"			
/db_xref="taxon:138677"			
/note="synonym:Chlamydia pneumoniae (strain J138)"			
144. .935			
/gene="CPj0267"			
144. .935			
/gene="CPj0267"			
144. .935			
/gene="CPj0267"			
/product="hypothetical protein"			
/protein_id="BAA98477.1"			
/db_xref="GI:8978641"			
/translation="MSISLNTKTNALLNQEPAPCLNAMDPKYNQDKTFACVTYLL			
VATLMTITGVIVLWAGSPGLSVLSTIIIGTSTYTGALFTIGLTKLTKSLAMT			
OYKRYQEVNKKRPEFSPKNDVNHKLTSCPLSDIESPSPASFPVSKRLACSG			
VAIVLGVTLIGAVSVFECTGYQLDALCVGAGTALFPVGLAGLRTSHSLINQGM			
YELIVYLSSALRENEYTKDQNRNINLYLFECCQKREKALLE"			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
/codon_start=1			
/transl_table=11			
/product="hypothetical protein"			
/protein_id="BAA98478.1"			
/db_xref="GI:8978642"			
/translation="MSQCSSTSTWEMKSRVPMKKMPPTPLSPISDEDFILAYEF			
FVLPTDPENAQANPEGSTPTPVNGLIDPLNLLGQPEQNNANNPSTSGSNPSTLPA			
PERLPETENSOEEQSQNNEDLIG"			
complement(1495. .2472)			
/gene="CPj0269"			
complement(1495. .2472)			
/gene="CPj0269"			
complement(1495. .2472)			
/gene="CPj0269"			
/codon_start=1			
/transl_table=11			
/product="dipeptidase"			
/protein_id="BAA98479.1"			
/db_xref="GI:8978643"			
/translation="MTIDMCDLLSHPHFCRKPAPVRCSPQLLSGVROQCAIFVW			
HSGEPNCDKQNSLFEPLNQYDGLGILSYEEENGSSQKSLISLISNAGALG			
DRPGLTLLAKLILHFKQPLAYLGLIWMKGDMRFGGGEAPKRLSNDKVLDDIME			
GVPIIDSHCSDKLIEDLIDYTDKLPNLAVLASHNPFSSVLDHNRNEDAKAKIVRR			
KVYIGLNVYRSVYDSDGLDEKHVLAHNRGLSTIVLGSDFPAENDEPNFEECCSS			
AEAHPLNOLHRIHSKGAESITLSSRAKFLKQIVQVNPKITDVAL"			
complement(2487. .3347)			
/gene="yW1C"			
complement(2487. .3347)			
/gene="yW1C"			
complement(2487. .3347)			
/gene="yW1C"			
/codon_start=1			
/transl_table=11			
/product="SuA5 superfamily-related protein"			
/protein_id="BAA98480.1"			
/db_xref="GI:8978644"			
/translation="MPDKKAQITFSLPEVMSAIIHOGKIVALTPTDVFGLSYASEA			
ERLVYLRKREPSKAPALVYNSIEDIENISCTPSPTAKLAOLPPGATLYVYHRNP			
RPKETLARIYDHSVYREIVDHCGTLIGTSANLSEFSAITAOETFRDFRDHDLCTF			
DKPCHSGLESTIVASDPLTYREGGLISVLENTAGTAKIFHRTSHAFSHKITYIVN			
KNEQQLSVFSLGSLDFKGVCEHPKKNFVRLREALKKTPSIYFIVDINTSDYPEL			
FPILSPYIE"			
3566. .4387			
/gene="CPj0271"			
3656. .4387			
/gene="CPj0271"			
/codon_start=1			
/transl_table=11			
/product="lysophospholipase esterase"			
/protein_id="BAA98481.1"			
/db_xref="GI:8978645"			
/translation="MTDYSFFRRKIGNIEATPCNPDPPIIILHGVGSLADNLTFEE			
PSICSPSKRLPWIPFNGTILPLENDFRSCAPFLNVLLOELSLVANGNGLOEYKY			
DELFDVDETPEKALEELIILNRYNELIIGFSGGLILTHIVISQNPYACALILFI			
ACARLPNCGMEGLKQCAQVPRLOSHTGHEDELTPYHLGAHLNDLLITKLNQFSPHIG			
GHEIPSVYEQKQVITPNNIDPARG"			
complement(4359. .5255)			
/gene="dnaX_2"			
complement(4359. .5255)			
/gene="dnaX_2"			
/codon_start=1			
/transl_table=11			
/product="DNA polymerase III gamma and tau"			
/protein_id="BAA98482.1"			
/db_xref="GI:8978646"			
/translation="MHLSEENGWALLRKVYHOEVPAILHGFLLPYLDOKKAQOLA			
SEILSSSPGSEHKVSKRHNIDYQFPEGGGRGRLSIDLPGRITKOIYISPEFNKTI			
YIITHEDRNTLALIASFLKVEEPKHAIVITLTKAKVORLPKTIISLSLIFERK			
ILCSKTFEYLFYVACETIPVTEVSQIILKESSEDKOVLROKVOREMEVLELEYRDY			
TNLGICAKALNYPHVEVEIQLDPLPLDKVLLIYESACRSILNSSSAVLEWVAIO			
IVSLQKEKEILSVSPGDLIS"			
complement(5246. .5866)			
/gene="tdk"			
complement(5246. .5866)			
/gene="tdk"			
complement(5246. .5866)			
/gene="tdk"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			
gene			
CDS			
957. .1343			
/gene="CPj0268"			



```

/codon_start=1
/transl_table=1
/product="thymidylate kinase"
/protein_id="BA98483.1"
/db_xref="GI:8978647"
/translation="MPLVIEGSGSGSSSLAKALDGLVADNRKVLITREPGCLIE
RLNDLLEPHLEISRCCELFELGSAHQHIEVILPALDGVIVICERHDSITVQ
GLAEGLADPVADLCKVSGPTPELPNVLIDLPADIGQRKHQKPFKEKKPLS
YHNRIRGEFLSLASADPSRYLVDARESLASLIDKVMLEHTQLGICT"
complement(5871..8375)
/gene="gyrA.1"
complement(5871..8375)
/gene="gyrA.1"
/codon_start=1
/transl_table=1
/product="DNA gyrase subunit A"
/protein_id="BA98484.1"
/db_xref="GI:8978648"
/translation="MKNKDEILYPKNLEEMKESLYRSMVITISRALPDIDGLKPS
ORVLYAMKQLSLSPGAKHRKCAKICGDSIGDHPHESVITPLNDEKHEVYVP
VDSQNGFSGIDGPPAMRYTEARLTHSAHLEIDLDKDDVADIVNDEKHEVYVP
SKFENLCSGSGIAGVAMATNIPPHNGELIEATLLDPAQSVDEILQVMPEDPP
TGGIICSEBIRSTYTRGKIKYRARLHVEENDKRESITITTEMVYNTKSLIEO
IANLVNKTLAGISVDKDGIRVYLEIKGSESEIINIKTKFTDVQVFGAM
LADKRLPTMSIHRMISAIHRKREYIRRTRELKKAETRAHVEGYLAKSLDIA
LAVTIRSGKKEHAKERIIESFGTEPOLALIELRYOLGLEAEIKQREYELNK
IAYKQVLSDEGLVKDIIINELQDLKHHVARTTEFPADIRFIDEDITFNSVYI
TISGDVYKRPMPKVEKORGGHVTGFPDKAGGLKAVSAFTDYLITFNEGO
CYMLKWOIPEGRRAKGRIPNLEIRGEELATLNKFNQNGEFLRLARKGVY
KRYSLDAFSNPKRGIRALEIDEGDELIAACHVYSEKWMFLTHLMAVFPHEKRA
PMGRITANGVSLKNEEDVVSQVITENOSVLIYCDQFGRSLVEDRETNRGV
GVSILINERNGVLAGIPLYTDHDSILMSQOQALINQDVRVMGRSTOGVRLVHL
KESDIALVSMKLSNENDEVLSGSEECSDTVSLR"
complement(8391..10808)
/gene="gyrB.1"
complement(8391..10808)
/gene="gyrB.1"
/codon_start=1
/transl_table=1
/product="DNA gyrase subunit B"
/protein_id="BA98485.1"
/db_xref="GI:8978649"
/translation="MDPKENYDASATVLEGLQAVREPMQITGDTGLHLVYE
VWNSIDEAMAGYCSRIDVRLIEDGGLVIVDNGRIPIEVEHESAGQGEVSALREV
PLTVLHAGGKDKDSYKSGGLHGVSCVVALSEKLVATVFKCKOMFESGIPYT
LQVYVSVDROGTEIVEYDPKIFSTCTPFRKGRIRLAFLEINRITITFEDDRY
SEPKVFEYEGGIOSFVSYLNOKKSELSFSEPIYICGRVADDEIEEALQNNQSG
ELVYSANNIPTROGSHLGTSTALTRVNTYIKAHNLAKNNKLATGDIIEGLTA
VISKVPNPOFEGQKOKLQNSDVSVAQVVEBALTFEENPQIARMLVDVFAVA
QAREAAKARELTLRKSALDSARLPGLIDCLEKDEPKCEMYIVEGDSAGSARQGD
RRFOALLPRLGKILNVEKARLQKIFONOEIGTIIAAGCIGADNEMLSLRYRRIIT
MTDADVDSHIFRLITFEYRHMTALENCVIAOPIKAVSKKRFRIEISEKED
SYLMIGTNESSILFKSTERELGALLESINVLIDESIINLEKKAIFESFLEY
KEGIGPLYLTIARATGMOGRVYLSDEKEEALAOETHNFKIITELKAVFVIDIQO
LKEGLDISSYLIPQKNEIYIGNDSFSCYTLIEVINYLNKRGKIEIQRKG
LGEWMDQIOWDITMNPQRTLLHVSLEKDAVEADHIFTMLGEEVPRPEIESHALSI
RINLNDI"
complement(10812..11159)
/gene="CpJ0276"
complement(10812..11159)
/gene="CpJ0276"
/codon_start=1
/transl_table=1
/product="Cris1 hypothetical protein"
/protein_id="BA98486.1"
/db_xref="GI:8978650"
/translation="MFLKRRKRGSGVQNKRTASPIKHAHYLHNTLQELQKTMARP
HAIADWQVFRDKRYGMSQAIQFRDHIILVKYNSLVALLKQTDNDLIMSLYQVA
SHVQIREIOFILG"

```

Query Match 23.1%; Score 649.4; DB 2; Length 300650;  
 Best Local Similarity 54.5%; Pred. No. 4.7e-145;  
 Matches 1548; Conservative 0; Mismatches 1231; Indels 60; Gaps 10;

```

QY 1 ATGAATGCAATTTTCCGTGTAGTCTCTTCGACATTCGATGTTTACTAGTTGT 60
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197343 ATGAATAATACCTTGACAAACCTCTGATCTCTTGACATCTTGACATCCCATCTATGT 197402
QY 61 TCCACTGTTTTCGTCGACACTGCTGAATAATATAGGCCCTCTGATAGCTTTGACGGA--- 117
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197403 AGCATTCGACACTTACGGGAGCATGCTCTTATACCCCTACAGATAGCTTGAATGGACCG 197462
QY 118 AGTACTACACAGGACCTATATACCTTAATAATAGACTACTGGAATGACTATACTCTG 177
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197463 GCGGCTCTACATTTTACCTCCAAATCTTACACAGATGCAATGGAACACTATATGTCTTA 197522
QY 178 ACAGAGATATTAATCTTTCGCAAAACCTTGGGATTCGACGCTTACGAAAGGTTGTTT 237
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197523 TCAGAAATGCTATATTAACGATGCTGGAAAGCAGCAGCATTAACAGCTGCTGCTTT 197582
QY 238 TCTGACACTACGGAATCTTTAAGCTTTGCCGTAAGGGGTACTACTCTTTTAAAT 297
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197583 ACAGAACTACGGGTGATCTGACATTTACTGAAAGGATGATCTATTTTCAATCAACAG 197642
QY 298 ATTAAGTCTATGTC---TGAAGGCGCAGCACTTCTGTACACATGATTAATCTGTGCG 354
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197643 GTAGATCGGCTTGAATGACAGAGCTGCGCAAGCACAACACTGCTGATTAAGCCCTTACA 197702
QY 355 CTAACAGATTTTTCGAGTCTTACTTCTTACCGGCCCCCATGCTGATATCAACACCCC 414
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197703 TTCACAGATTTTTCACCTTCTCTTCACTGACACTC-----CTGGAACATACAGTTGCT 197756
QY 415 TCAGAAAGATGCACTTAATATGAGAGGAGTCTTACATTTGATTAACATGCAACTAT 474
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197757 TCAGAAAGATGCACTTAATATGAGAGGAGTCTTACATTTGATTAACATGCAACTAT 197816
QY 475 TTATTTAAACAAGATCTGTGAGGAA-----ATGCGCGAGCATTTTACCAAG 525
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197817 CTTTTCACCAAAACGCTCCAAATGACCTATATACAAATGCGAGGATACCGGCAAAA 197876
QY 526 AATCTTCTTTGAAAACAGCAGCAGGATTCATTTCTTTGAAAGGAATTAATCAGCGCA 585
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197877 ACTCTTCTATATTCGGGAATACGCTCTTATTAACCTTCACTAGTAATVAGGCAAAA--- 197933
QY 586 ACAGGAAAAAAGTGGGCTATTTTGCTACTGTGACTGATATTTACAAATTAATACG 645
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197934 -----AATVAGGTGAGCGCATATAGCTTCGCGCTGCAATATTTTCAGAAACCC 197987
QY 646 GCTCTACCCCTCTCTCGAACATATGCTGTAAGCTCAGGTGAGGTATTAATAGACA 705
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197988 GGCACGTAGCTTTATGATTAATTAAGAGAAACTGGGGGTGGGCTTGGCTTTGAA 198047
QY 706 GGAACGTACAAATTACAGGGAATAGCTCTCTGTAATTTCTGAAAATAGTGACAGCG 765
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 198048 GCCAGCTCCTCGATTAATCAAAATAGCTCCCTTTCTCTGGAACACATGCAACGAT 198107
QY 766 ACCGACAGAAATGACGAGCTCTTTGAGAGA-----TGCAGATTTAC 810
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 198108 GCTCAGCAAGGCGGGGCGCATATATGTAAGAAAACAGAGAGACTCTTACTACT 198167
QY 811 ATATCTGGGAATCAGAGTACTTCTTACAGGAACCAAGCTGTAGTAATGGCGGAGCC 870
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 198168 ATCTCTGAAATTAAGTCTGACCTTGGCGGAGACACTCTTCAATGATCAAGGCGGAGCA 198227
QY 871 ATTTAGCTAAGAACTTACACTGCTTCCGGGGGGGGGGGATCTCTCTTTTTCAC 930
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 198228 ATCTGCGCCATGCTGATAGATCTTCCGCTGCTGGCCC-----TACCCTATTTTCAAT 198281
QY 931 AATATAGTCAAGTACCACTGACAGATAGTGTGACCAATTTTATATCTGGACGCTGGA 990
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 198282 AATGATGCGGGAACACAGCTGACGAGGCGGCGCTATGCAATTTGCCGACTGGA 198341
QY 991 GAGTGTAGTCTTACAGGAGGAGGAGGAGTACCTTCAATGGAATGCCATTTGCA 1050
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 198342 TCTTTAAGTCTCTGCAATCAAGGAGACATCAGTTCCTTGGCAATCTTAACTCA 198401

```



OY	1051	ACACACCCAAACTACAAAAGAAATTCATATGACATATAGATCTACAGCAAAATGACAG	1110
OY	1051	ACACACCCAAACTACAAAAGAAATTCATATGACATATAGATCTACAGCAAAATGACAG	1110
Db	198402	ACCTCCGGCCAAATATGACAGGAATGCTATCTACCTGGGATGCTGCACAAAATATACG	1984
OY	1111	AATTACGTGCATATATCTGGGCATATACATCTTTTCTACAGATCCGATTAAGCTAATACG	1170
Db	198462	AACTTAAGGCGAGCCCAAGGCCAATCTATCTATTTTCTATGATCCGATTCATACAC	1985
OY	1171	GCTGGCAATTCACAGATACTTTTAAATCTCAATTAAGCTGATGCAAGTAATAGTACAGAT	1230
Db	198522	ACAGAGG---CTTCAGACGTTCTGACCATCAACCAACGAGATAGCACTCGCTTTAGAT	1985
OY	1231	TATAGTGGGTGCAATGTTTTTTTCTGGTGAAGAGCTCTGAAGATGAAGCAAAAGTTGCA	1280
Db	198579	TATTCAGGAAGATGTATTTTCTGGGAAAAGCTCTGCGAGATGAAGCAAAAGCTGCT	1986
OY	1291	GACAACTCACTCTACGCTGGAAGCAGCCTGTACCTTAATCGCAGAAATTAATGTAAT	1350
Db	198639	GATATCTTCAATCTATATTAATTAAGCAACCATTTGGCTCTTAGCTCTGGAACTTTAGCACTC	1986
OY	1351	AAACGTGTGTCACTCTCGATPACGAAAGGCTTTACTCAGACCGCGGGTTCCTGTATAT	1410
Db	198699	AAAGGAATGTGCGATGTAGATGTCAATGTGTTTCACACAGACTGAAGGCTACACTCTCTC	1987
OY	1411	ATGATATCGGGCACACACGTTTAAAGCAAGTACAGAGAGGTCACTTTAACAGTCTTTC	1470
Db	198759	ATTCACACAGAAACAAAGCTCAAAAGCAGATGACTGAAGTATAGCTTACCAAACTGTCTC	1988
OY	1471	ATTCCTGTAGACTCTTTAGGCGGAGGTAGAAAGTGTATATGTGCTTCTGCGACGAAT	1530
Db	198819	GTTATCTCTTCTGCTTACGAGGGAATTAAGATGTGTCATTGAAACAGCAGAGCCAC	1988
OY	1531	AAAATGTAGCCCTTATGTGTCCGATCTCTCTTTGTGATTAACCAAGGAATGCTATGAA	1590
Db	198879	AAAATTAATCTCTAACCTCTCTCTGTTTTCCAAAGATAGTAGGGCAATTTTATGAA	1989
OY	1551	AATPACAGCTTAGGAAAACACTCAAGACTTTTCAATTTGTGCACTCTCTGCTGGATCT	1650
Db	198939	AGCCATACGATTAACCAAGCCTTCACGACGCTTTGTGTGTATCTACCTGCTACTGCT	1989
OY	1651	GCAACACTACAGATGTGTCCAGGGTTCCTACAGTACGAACATCTAGCAGCTATGGGAT	1710
Db	198999	GCTAGCCATTTTATATCGATGAGGTGCTTCTCTCCAGTACAAATCTCAGAACTCAT	1990
OY	1711	CAAGGTACTTGGGGAATGACTTTGGTGTGATATACGCAAGCACTCAAAAGCTAAGACA	1770
Db	199059	TACGGATATCAGGA---CATTTGGAAAGCCACTTGGGCGAGACATCAACTGCGCAAAATCA	1991
OY	1771	GCGACATTAAGCTTGGACCAATPACAGGCTACTCTCGAATCTCGAGCGTCAAGGACCTTTA	1830
Db	199116	GGAATCTATGACTTGGGAATCTACGAGCTCAACCCATTAATCTGACGTGATGCTTCGTA	1991
OY	1831	GTTCTTATATAGCTTTGGGATCTTTTTCAGACATTCGCAAGGATTCAGAGGTGCTATAG	1890
Db	199176	GTTCCGATTAATTAAGGGATCTTTTACTGCAATTCGACTCTACACAGATCATGACA	1992
OY	1891	AGAATGCTTTGACTCTTTGTCAGATCGAGGCTCTGGGCTGCGGGAGTGGCCAAATTC	1950
Db	199236	TCTTAACGCAATATCTATCTATACGACAGAGACHTGTGGCATCAGGAACCTGCGAAATTC	1992
OY	1951	TTAGATTAAGATTAAGAAAGGGAACCAACGCAATACCGTCATTAATCTGGTGAATAGCT	2010
Db	199296	TTTCATTAAGATTAATACAGGAATCAACCAAGCATTCGCAATTAAGGCTAGGGCTATATTT	1993
OY	2011	ATCGGAGTGCAGGCAAACTTGTGTGAACCTTAATAGCTTTGCTTTGGCAACTC	2070
Db	199356	GTTGGAGGAAGTCTGTAAAGTTTTTCTGAAAATCTTCAGTGTAGCTTTCTGCGACCTC	1994
OY	2071	TTTGTGAGCGATAAAGTTTCTTAGTCGCTTAATAATCAATAGCAATCTATAGCAGAGGCC	2130
Db	199416	TTTCGTAAAGATTAAGACCTGTTTATAGTTGAAATTAACCTCTCATATCAATTTTAGCGTCG	1994
OY	2131	TTTCATATCCACACATTAACGAATGTAGTGGTTCAATAGTTGTCTCTTGTGATTAACCT	2190

Db 199476	CTATACCTGCAACATCCGAGCATTCCTTAGAGGAGACTTCCCATGCCCTCATTTGGAAAGTATC	199535
QY 2191	CGTGCCCTTTGGAGTCATTAACCCCTCGTTTATGAAAGGAGCTCGCTATTATGACCAAGTC	2250
Db 199536	AACGACATGCTGAAGAAAGATATTTCTCTCATTTTGAATGCCAGCGTAACGTCACGTCACACT	199595
QY 2251	AGTATATATCTGAAGACAAAGATATACTCGCGTATCTCTAGGTGAAGGTCTTTGGGGAAAT	2310
Db 199596	AAAAATGATATGATGATACACGCGTAACTACTCTCATCTTAACCTCAAGAGCTCTTGAGCAAT	199655
QY 2311	AATGCTTTTAACATGATGTGGAGCTTCCT--TCGATCTTATCCGTGAATACCTGCAT	2367
Db 199656	AACCTGSGGCGCTTAGAGACTGGAGAGATCTCTGCGCTATATCTCCCTAAAGAAACACCG	199715
QY 2368	TGTTTGTATACCTATGCTCATACATCAACATGAACTGAACTGACCTATATATAGTCAGAGACGC	2427
Db 199716	TTCTTCAGGAGATATTTCCCTCTTAAAGTTCAGGACAGTCACAGCCCAACAAAC	199775
QY 2428	TTTTCGGAAGAAAGGTACAGAAAGAAAGCTTTTGTATACAGCAACCTCTTCATTTATCT	2487
Db 199776	TTTAAAGAGATGGCGCGCTGAAGCCCGCTGCTTGTATGATATGAGACCTAGGAAGCTGCT	199835
QY 2488	TTGCGTATAGGGGTGAAGTTTGAAGAGTCCTGATTTGATGACTTTCTATGATCTG	2547
Db 199836	ATCCCTGTGCGCATTCGCTTAGAAAAATCTCCGAAGATGAAAAAATAATTTCGAGATT	199895
QY 2548	ACTTTATCCATATGTTCTGATCTTATCCGCAATGATCCCAATGACATACAGCACTTGA	2607
Db 199896	TCTCAGGCTACATTTGGTATGTATACGTAAAAAATCCCGCTTCGCTACTCTCTAATG	199955
QY 2608	ATCAGCGGAGGCTCTGTGGAAACTTATGCCAATTACTTACAGACAGGCGCTTGCAAGTC	2667
Db 199956	GTCAGTGAAGGCTCTTGGACTTCCCTATGTAAAAAACCTGCAGACAGGCTCTTAGCA	200015
QY 2668	CGTGACGACGACTACATACGCGCTCTGCTCCATGTATGTTAAAGGCTCGGCGCAGTTTGTCTT	2727
Db 200016	AGTGTGGAAACCATCTGACCTCTCTCCCTCATGTAAACTCTCTGSGGGAAGCTGCTTAT	200075
QY 2728	GAAGTTCGTGATCCCTCAGCGATTATATATGTAGATCTTGGGGGTAAAGTTCAAATTCAG	2787
Db 200076	GAGCTTCGCGCTCAGCACACATCTACAAGTGTAGATTGTGGGCTAAAGATACATCTTAG	200135
QY 2788	GAGGCTCTCTCATGTCTCA 2806	
Db 200136	TTCTACTTTCTCTCCCTAA 200154	

RESULT	14
LOCUS	CPI133034
DEFINITION	CPI133034 17280 bp DNA BCT 27-FEB-2001 Chlamydia pneumoniae GGAI-protein family cluster A, genes omp12 (partial), omp11, omp10, omp5, omp4, omp13 and omp14 (partial) and ORF1, strain VR1310.
ACCESSION	AU133034
VERSION	AU133034.1 GI:4455881
KEYWORDS	hypothetical protein; omp10 gene; omp11 gene; omp12 gene; omp13 gene; omp14 gene; omp4 gene; omp5 gene; ORF1; outer membrane protein; outer membrane protein 11; outer membrane protein 12; outer membrane protein 14; outer membrane protein 4; outer membrane protein 5. Chlamydophila pneumoniae. Chlamydophila pneumoniae Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila. 1 (bases 1 to 17280) Daugaard,J., Hjerno,K., Knudsen,K., Madsen,A.S., Christiansen,G. and Birkelund,S. Unpublished 2 (bases 1 to 17280) Boesen,T. Direct Submission Submitted (21-JUN-1999) Boesen T., Department of Medical Microbiology and Immunology, University of Aarhus, The Bartholin
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
JOURNAL	
TITLE	
REFERENCE	
AUTHORS	
JOURNAL	

Building, DK-8000 Aarhus, DENMARK  
Location/Qualifiers  
1..17280  
/organism="Chlamydomonas reinhardtii"  
/strain="VR1310"  
/db\_xref="taxon:83558"  
1..841  
/gene="omp12"  
<1..841  
/gene="omp12"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein 12"  
/protein\_id="CAB37067.1"  
/db\_xref="GI:4455882"  
/db\_xref="SPTREMBL:Q9Z504"  
/translation="EDNIRYRHNSGYALITATTPAEDOLTFAFCOLPARDNRHTTG  
KMHGDTYGASLYFHHTEGLDPAIDNFMKGATRAWVLEISQIPLSEDAKFSYLAHD  
NHHKTYTDSNIIKGSWRNDAFCADLGASLPVIVSYLLKEVEPFVKVQYIYAHQOD  
FYERHAGRAFNKSELINVEPLIGVTERDSKSEKGTIGYLLKLYILDAYRNPICOTS  
LIASDANMMAVGTILNARQFSVRANHPQVNPHEITGCAFEBVRSSRNNTNLGSK  
PCP"  
885..967  
/gene="omp11"  
885..3677  
/gene="omp11"  
885..3677  
/gene="omp11"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein 11"  
/protein\_id="CAB37068.1"  
/db\_xref="GI:4455883"  
/db\_xref="SPTREMBL:Q9Z393"  
/translation="MKIPLKLLISTLVLPILISATYGAADSLSEPTSDGAGGST  
FTPKSTDANGTIVLVSGNVIYINDAGRTALTGCCFETTGDLFTGKYSFSFNTVD  
AGSNAGAASTTADKALFTGFSNLSFIAAGTIVASGKSTLSAGALNTDNTILF  
SONVNNANNNGAIIITKTLISGNTSITFTSMARKKIGAIYSAASITSGNTGQL  
VPMNKGCTGGALGPEASSITGNSISLFGSNCTAPDAKGAICYCEKGERPGLT  
SGKSLTFPENSSTYOGGATCAHGLDLSAAGPTLFSNNRCGNTAKGGAIALADSS  
LISANOGDITFLGNTLTSTAPTSTRNALYLGSSAKITMLRAQGOSIIFYDIAIN  
TGASADLTINOPNSNPLDYSQFIYVSGKLSADEKADNFTSILKOPALASGLT  
ALGNVELDVNGFOTEGSTLLMOPGTAKADEALISLTLLVDLSALEGNKSVSIT  
AGANKTITLSPLEFODSGNFESEHTINQATOPLVFTAAASADITYIDALTSTV  
OTQPHGYOGHMEATWADISTAKSGTMTWTGYNPNPERASVYPDLSMASTDIR  
TLQOIMTSQANSIYQGRGLASGTANFPHKDSGTNAPFRKSGIYVGSASPEFSN  
IFSAFCQLTGKDKDLFIVENTSHNTYLAHLQHRALVGLPMPFSGSIIDMLKDL  
ILNAQLSYSTKNDMDRTYSYPAQGSWTNNSGALVGLSLALYLPKEAFPGQYFP  
FLKFOAVYSKQONKEGAGARAFDDGDLVNCSPVGIKLEKISEDEKNFETISLAVI  
GDVYRKPRRSRTSLMVSFASGWSITCKMLARQAFILASGSHLTLSPHVELSGEAAVELR  
GSAHYVDCGLGRYSF"  
3827..3907  
/gene="omp10"  
3827..6613  
/gene="omp10"  
3827..6613  
/gene="omp10"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein"  
/protein\_id="CAB37069.1"  
/db\_xref="GI:4455884"  
/db\_xref="SPTREMBL:Q9Z398"  
/translation="MKSSLHWFLISSLALPLSLNFSARAAVEINIGPTNSGPGCT  
YYPQATYNADGTTIYNLTGVSITNAGSPALPASCARETGMNLSFGHGYOFLQNI  
DAGANCTFTANKLILSFSGSYLSLIQTTNATGGAISTGACSIOGNSCYFQO  
NFNDNGALOGSSISLSLNPILTPAKNKATOKGALVSGGITINNTLNASPESNT  
AANNGAIYTBASSPFISSNKAISFINKSVATATGCAIYCSSTSAKPKVLITSDNE  
LNTIGNTAITSGAIYTDNLVSSGGTLEKNNASIDTAPLGAIALIASGSLISA  
LGDLITFECHTIVKASSSOTTTTRNSINIGNTAKIVQLRASOGNTIYFPYDITTSIT

ALSALDNLNGPDLAGNPAYQGTIVFSGERLSEAEAAEDNLKSTIOQPLTLAGQLS  
LKSQVTLVAKFSQSPGSLTLMADGTLFETADGTTINNLVANDSLKETKAKLQAO  
ASQVTLGSLTLDVPSGNVYEDSNVNNPOVSGCLTTPADQANIHITDLAADPLEAO  
PIHMGYOGNNALSMQEDPATKSKAATITKRTKGNNPBERGTLVANTLWGSYDVS  
IOOLVATRVKQSOETRGICWEGSINFPFKSTKINKFRIISAGIYVGAATTLASDL  
ITAFVQLFGKDRDHFINKNRASVYASLHLOHATLSSSLKLYLFGSSSEQVPLD  
AOISYISKNTMTKTYTOAPRGSSWVNDCALELASLPHLASHGLHAAPFPK  
VEASYIHODSEKERNTLVRSFDSGLINVSVPGLTFEERFSRERASVYATVYVAD  
VYRNKPCCTTALLINNTSMKTTGTNLSRQAGIGAGIFVAFSPLEVTSMLSMERGS  
SRSYNADLGKRPQ"  
complement(6954..8060)  
/note="ORF1"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAB37070.1"  
/db\_xref="GI:4455885"  
/db\_xref="SPTREMBL:Q9Z897"  
/translation="MSPFQOEQCHFDVQSFPLRPESLTPRARSDFEGRVIEQOMRV  
EDAIKMLIKQTEAGLIFPTDGEFRFVYSDPDMQFHVGRDRRSDNDEIYVLMKD  
KISYKSPLEHFEFVLEFPEKNAKATLPSPQSFHEKMTFAPNLKNTKRPVPTNOE  
LIDIVYVROVIOIDLVAACGRNLOLDNCAWCRILDIRASWTVGSDHDELOELBOF  
LMIHNLNMRKRPEDLPYSLHVCQDQYAEFRSRAVDIEPELFAKDYVSXYHYVAL  
DKRTSGGAEPVATVSGEHVCLGLISNSHCIEDRQVVSRIYEASVIFLERLSIP  
QCGFASCEGDHMRTEEQMKRIAFVKELAEIENG"  
complement(8586..11372)  
/gene="omp5"  
complement(8586..11372)  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein 5"  
/protein\_id="CAB37071.1"  
/db\_xref="GI:4455886"  
/db\_xref="SPTREMBL:Q9RB65"  
/translation="MKQSPWLVSLTFLACFSCTYFATTAENIGRSDSPDSSTNG  
TYPKNTTIDITLTDITLQNGDAAALIKGCFSDTSLSPAGKISLSPUNIS  
SABEALASVTTDKMLSLTGPSLFLAPASVYITTPSGKAVVCGGDLTPDNNGTILF  
KQVETLTPKFTDPTAGSSVIMDAGTLAKSTEEVTLGLSIPVDSLGEKGVYIAASA  
ASKNVALSGTILLIDNONGAYENHDLKTDSEFVQVLSAGTATTTDVPVAVAP  
HYGQGTGKMGWDDTASTPKTKATLATNTGGLPPEBOGLDPLVSLGSSSDIOA  
IOGVILERSALITCSRDGFMAGVAFNLDKQKBEKKRYHRKSGVYIAGAAQONCELU  
ISPAFCQLRSDSDDELAKNHTPYAGATPIQHTDESGRTICGLLKLKLPSSMKHPI  
LEGLATSHVSNDKTYTAYPEYKSGWGNNAENMLGASHSYPTLHCFDYAPYI  
KLNLTYLRQSFSEKGTGERSFDSNLEFNLSLPGVFEKFSQCNDSYDLTLISVPD  
LIRNDPCTTALVYISGASWETVANNLRAQALQVRAQSHVAFSPFVTLGQFVEBRS  
SRIYVNDLGKRPQ"  
complement(11313..11372)  
/gene="omp5"  
11535..14321  
/gene="omp4"  
11535..11588  
/gene="omp4"  
11535..14321  
/gene="omp4"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein 4"  
/protein\_id="CAB37072.1"  
/db\_xref="GI:4455887"  
/db\_xref="SPTREMBL:Q86164"  
/translation="MKTSLPWLVSVAFLSCLHOSLANEELLSPDSSFGNIDSGTF  
TPKTSATYSLTDVDFEYEPGKCTPLSDSCFKOTNTNLFLGHSHTFEDIGACHA  
GAASATYANKNLIFSGSLSPDSSTVYTTGCGTSSAGVNLNIRKLVVAGNS  
TAGGAIKAGNLFLLTGSGDALPSSNSSSTKGAIAATTAGARIANNMGVYRFLSNAS  
TSGGALIDDECTSLSNKRPYFEGNAKTTGGAICHTKAGSGPELLIISNKTILPASN  
VAETSGAIIHAKRALVSSGGTFELKNNVSAIPKGAISIDSGELSSAEETGNTF  
VRNTLTVTGSTDTPKRMAINIGSNKRFETELRAKNHTIFEDYDITSEGTSSDLKINN

GSAGALNPVOGTLIFSGETLLADELKVADNLKSEFQPSVLSGGKLLLOKQTLTESTS  
ESQAGSLIGMDSGTTLTSTAGSTLITINDSLIGLKPVSLLTAAGAKNKVYSGSK  
LNLIDIGENITYESHMSHSDLPFLKLTITVADNDVRLINDSLIPVREDBNSENSEBGRG  
QMNVMYTDITATNTKEATATWTKTGPSPERKSALVCNLTLMGFTDTRSLQQLVEIG  
ATGHEKQFQWVSSMTNGLHTGDENKGRHRTSGGVYIGSATPKDDLPFAFPL  
FARKDCEFIANNRSRTYGTLLFKHSHRTLOPOMYLRGRKFSLSALEKPREPLAL  
DVQVFSHSDNRMETHTYSLESEGSSWNCIAGSLDLPFVLSNHPLEKFEIPDM  
KVEWVYVSONSPFESSSDGRGFSIGRLNLSIPGAKFVGQDSDYTDYDLSGPFVSD  
VVRNNPOSTATVMSPSMKIRGNLSROAFLRGSNNYYNSNCELFGHYAMELRGS  
SRMYNDVGTCKLRF\*

s19\_peptide  
14579..14644  
/gene="omp13"  
14579..16123  
/gene="omp13"  
14579..16123  
/gene="omp13"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAB37073.1"  
/db\_xref="GI:4455888"  
/db\_xref="SPTREMBL:0923D6"  
/translation="MTLRNPLTCSALFLALPAAQVYVLIHESDYGALNINKSLEPK  
ITCYPEGTSYIFLDVRIISNVKHOEDGAFINRSGNLFPMGNRCNFTFHNLTGEGF  
AALSINRGDITLILSNFSYLAFTSAPLIPQOGAIFYSLGSMIENSEEVEFCGYSM  
SGAITYPYLIGSKASRPSVNLGNRLVLRDMVSOVGGAISTHNLTLTRGSCOPE  
NNHAYHDVNSNGAIAIAPRGSISISYKSGDILFKGTTASQDGTTHNSIHLDSGAF  
KNLRAVSEGVYTFDPISSHESKITDVLINAPBEGKTVEGYSLSGLCDHDEVCE  
NLTSTLIQDVTLLAGGTLISDGYTLQSHLSKQESSTLTLSPGTLTLCSDARDQNH  
ILIEDDNEFVPRIRADKDALVLEKLYAFAYWYVDFPQKKEFTIPLLELGP  
SPDSLIGETTLERTQVTTENDAVRCHWSLSMEBYPISLDKRRITPPTKTVFLWNP  
EITSTP"

Query Match 23.08; Score 647.8; DB 2; Length 17280;  
Best Local Similarity 54.58; Pred. No. 1e-144;  
Matches 1547; Conservative 0; Mismatches 1232; Indels 60; Gaps 10;

QY 1 ATGAATGCAATTTTCTGCTAGTGTCTCTCGACATGGCATGTTTACTAGTGT 60  
DB 885 ATGAAAAACCCCTGCACAACTCCTGATCTCTCGACTCTTGACATCCCATTCAT 944  
QY 61 TCCACTGTTTTGCTGCACATGCTGAAAATATAGGCCCTCTGATAGCTTTGACGGA 117  
DB 945 AGCATTCGACATTCAGGAGAGATGCTCTTATCCCTACAGATAGCTTGTGATGACG 1004  
QY 118 AGTACTACACAGGCACCTATATCTCTAAATAGACTACTGGAATAGACTATATCTG 177  
DB 1005 GCGGCTCTACATTTACCCAAAATCTACACAGATGCAATGGAACGACTATGTCTTA 1064  
QY 178 ACAGAGATATACCTCTGCAAAACCTTGGGATTCGCGAGCTTTAAGCAAGGTTGTTT 237  
DB 1065 TCAGGAATATGCTATATTAACGATGCTGGAAGGACACAGCATTAACGGCTGCTCTT 1124  
QY 238 TCTGACACTACGATCTTTAAGCTTTCGCGTAAGGGGATCTGACTCTTTTAAAT 297  
DB 1125 ACAGAACTACGGGTGATCTGCAATTTACTGGAAGGATCTCATTTCTTCAACACG 1184  
QY 298 ATTAACTCTAGTC--TGAAGCGCAGCACTTCTGTGTAACACTGATAAATCTGTG 354  
DB 1185 GTAGATGCGGCTTGAATGCGAGAGCTGCGCAAGCAACATGGAATAAGCCCTAACA 1244  
QY 355 CTACAGAGATTTGAGATCTACTTCTTAGCGGGCCCATCATGCGTATACAAACCCC 414  
DB 1245 TTCACAGGATTTTAACTCTTCTCATTTGACGCTC-----CTGGAACCTACAGTTGCT 1298  
QY 415 TCAGGAAAAAGTGCAGTTAATGTGAGGGGATCTTACATTTGATTAACATGAGACTAT 474  
DB 1299 TCAGGAAAAAATCTATTAAAGTTGCGAGGACCTTAATCTTACCGAATAAGGAACAT 1338  
QY 475 TTATTTAAACAAGATTACTGTAGAA-----AATGCGGAGACCATTTTCAACAAG 525  
DB 1359 CTCTTTACCAAAAGCTCCCAATGAAGCTAATAACATGCGGAGCATCACCAAAA 1418  
QY 526 AATCTCTTTGAAAAACAGACGAGATCGATTTCTTTTGAAGGAATAAATCGACGCCA 585

DB 1419 ACTCTTTCTATTCTGGGAATACCCCTCTTAATACCTTCACTAGTAATAAGGCAAAA---- 1475  
QY 586 ACAGGAAAAAAGTGGGCTATTGTCCTCTCTGCTACTGTAGATATTACAATAATATACG 645  
DB 1476 -----AATTAAGGCGGAGCATATATAGCTCTGGCGTGAAGATTTCAGAAACACC 1529  
QY 646 GCTCTACCCCTCTCTGCAACATATGCTGAAGTGTGAGGTGAGGAGCTAATAATAGACA 705  
DB 1530 GCCCAGTTAGTCTTATTAATAATTAAGAGAACACTGGGGCGGGGCTCTGGGCTTTGAA 1589  
QY 706 GGAACCTGTACAAATTAACAGGGAATACGCTCTGTTGATTTTGTGAANAATGTGACACG 765  
DB 1590 GCCAGCTCCTGATTATTCACAAATAAGCTCCCTTTCTCTGGAACACATGCAACATAT 1649  
QY 766 ACCCGAGAAATGAGAGAGCTCTTCTGAGA-----TGCCGATTTAC 810  
DB 1650 GCTGACGCAAGGCGGGGCCATTTATGTGAAAAAACAGAGAGACTCTACTCTTACT 1709  
QY 811 ATATCTGGGAATCAGAGCTTAATCTTCTCAGAAAACCAAGCTGTATGCTAATGGCGGAGCC 870  
DB 1710 ATCTCTGGAATTAAGCTCAGCTACCTTCCGCGAGAACTCTTACATCAAGCGGAGCA 1769  
QY 871 ATTTATGCTGAAGAAGCTTACACTGCTCCGGGGGGGGGGGATCTCTCTTTCTAAC 930  
DB 1770 ATCTGTGCCCATGCTGATGATCTTCCGCTCTGCCCC-----TACCTATTTTCAAT 1823  
QY 931 AATATAGTCAAGGTACCACTGACAGATATGATGAGGACATTTCTATATCTGGACCTGGA 990  
DB 1824 AATAGATCGGGAACACAGCTGACAGGCAAGGGCGGCGCTATGCAATTCGCGACTCGA 1883  
QY 991 GAGTCTACTCTTTCAGCAGAAAGCAGGGGACATTAACCTTCATGGGAATGCCATTTGCTCA 1050  
DB 1884 TCTTAACTCTCTCTGCAAAACAGAGACATCACTGCTCTGCAACATCTTAACCTCA 1943  
QY 1051 ACTACACCAACAACTACAAAAGAAATCTATTGACATAGATGATCTACCAAGATCAAG 1110  
DB 1944 ACCTCGGGCAACATGCAAGGATGCTATACCTGGATGCTGACAAAATATACG 2003  
QY 1111 AATTTAGCTGAATATCTGCGCATAGCATCTTTTCTAGCATCCGATTTACTGCTAATACG 1170  
DB 2004 AACTTAAGGAGCGCCCAAGGCGCAATCTATCTAATTTCTATGATTCGATGATCAACCC 2063  
QY 1171 GCTGGGATTTCTACAGATACCTTAAATCTCAATAAGCTGTATGCAAGTAAATAGTACAT 1230  
DB 2064 ACAGAG--CTTGACAGCTTCTGACATCAACCAACCGGATAGCACTCGCTTATGAT 2120  
QY 1231 TATAGTGGGTGATGTTTTTCTGCTGTAAGAAAGCTCTGTAAGATGAAGCAAAAGTTGCA 1290  
DB 2121 TATTCAGGAAGATGTTATTTCTGAGGAAAAAGCTCTCTGCAAGATGAAGCAAGCTCT 2180  
QY 1291 GACAACCTCACTTACGCTGAGAGAGCGCTTAACCTTAACCTACGAGAAATTAATCTT 1350  
DB 2181 GATTAACCTCACTATATTAAGCAACCATTTGGCTGCTGAGCTGCAACCTTACACTTC 2240  
QY 1351 AAAGGTGTGCTACTCTGATACGAAAGGCTTTACTGAGCCGGGGTCTCTGTATTT 1410  
DB 2241 AAAGGAATGTGAGTTAGATGATCATGTCTTCAACAGACTGAAGGCTTCAACTCTTC 2300  
QY 1411 ATGATGCGGCAACAGCTTAAAGCAATGACAGAGGAGGACACTTAAACAGGCTTTCC 1470  
DB 2301 ATGCAACAGGAACAAAGCTTAAGAGCATATCTGAAGCTATACGTTTACCAAACTTCTC 2360  
QY 1471 ATTCCTGTAGACTCTTAAAGGAGGGAAGAAATGTAATGTGCTGCTTGCAGCAAGT 1530  
DB 2361 GTTATCTTTCTGCTTAAGAGGAATTAAGAGTGTGCATTTGAAGACGAGGAGCAAC 2420  
QY 1531 AAAATGTAGCCCTTATGTTGCTCGATTTCTTTTGATTAACCAAGGAATGCTTATATA 1590  
DB 2421 AAAATTAATCACTTAACCTCTCTCTGTTTTCAGAGATATAGCGGCAATTTTATATA 2480  
QY 1591 AATCAGCACTTAGGAAAAACCAAGACTTTTCATTTGAGAGCTCTGCTGCTGGTACT 1650

Db 2481 AGCCATACGATTAACCAAGCCCTTCAGCAGCCTTTGGTGATATTCAGTCGCTACTGCT 2540  
Qy 1651 GCAACACATACAGATGTTCCAGCGGCTTCCTACAGTACAGCACTCTACGACATGGTAT 1710  
Db 2541 GCTACCGATATTTATATGATGATGCTTCTCAGTTCTCCAGTACAACTCCAGAACCTCAT 2600  
Qy 1711 CAAGTACTTGGGGAATGATCTGGGTGATGATACCGAAGCACTCCAAAGATAAGACA 1770  
Db 2601 TACGGGTATCAGGGA---CATTTGGGAAGCCACTTGGGACAGACATCACTCAAGTCAAAATCA 2657  
Qy 1771 GCGCATATAGCTTGACCAATACAGCTACCTCCGATCCCTGAGCCGTCAGAGACTTTA 1830  
Db 2658 GGAATATATAGCTTGGTACTACAGGCTACCAACCTTAATCCCTGAGCGTACAGGCTTCGTA 2717  
Qy 1831 GTTCCTAATACCTTTGGGGAATCTTTTCAGACATCAAGGATCAAGTGTCTATAG 1890  
Db 2718 GTTCCGATCTATTTATGATGATGCTTCTTACGATTCGACATCTCAGACATCATGACA 2777  
Qy 1891 AGAGGTCTTGTACTCTTTGTTCAGATGAGGCTTCTGGGCTGGGAGTGGCCAAATTC 1950  
Db 2778 TCTCAGAGCAATAGTATCTATCAGCAAGGAGCTGCGCATCAGAACTGGCAATTTTC 2837  
Qy 1951 TTAGTAAAGATTAAGAAAGGGAAGGAAAGCAATTCGCTCAATTAATCTGGTGATATGCT 2010  
Db 2838 TTCCATTAAGATTAATTCAGAACTACCAAGCAATTCGCAATTAAGGCTACGCTATATTT 2897  
Qy 2011 ATCGAGGTGACGCGCAACTGTTCTGAAACTTAATTAAGCTTTCCTTTGCGCACTC 2070  
Db 2898 GTTGGAGAGATGCTGAGATTTTCTGAAATATCTCAGATGATGCTTTCGCGCAGCTC 2957  
Qy 2071 TTTGGTACGATTAAGATTTCTTACGCTTAAATCACTACTGATACCTATGACGAGCC 2130  
Db 2958 TTTCGTTAAAGATTAAGACCTTTATAGTTAAATTAACCTCTCAATTAATTAAGCTCG 3017  
Qy 2131 TTCTATATCAACACATACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2190  
Db 3018 CTATACCTGCAACATCGACATCTCTAGAGAGACTTCCCATGCTTCATTTGAAATATC 3077  
Qy 2191 CCGGCTCTTGGAGATCAATMAACCCCTGTTTAAAGAGGAGCTGCTTATAGCCAGCTC 2250  
Db 3078 ACCGATATGCTGAAGATATTCCTCTTAAAGTCCAGGAGCTAGCTAGCACTACACT 3137  
Qy 2251 AGTATATGCTGAAGCAAAATATCTGCTATCTGATGAGGAAAGCTTCTTGGGGGAT 2310  
Db 3138 AAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3197  
Qy 2311 AATGCTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2367  
Db 3198 AACTCTGGGCTCTAGAGCTGAGAGATCTCTGCTCTATATCTCCCTAAAGAAAGCACTG 3257  
Qy 2368 TGTTTGATACCTATGCTCATATCATCAATCAATCAATCAATCAATCAATCAATCAATCA 2427  
Db 3258 TTCTTCCAGGAGATATTCCTCTTAAAGTCCAGGAGCTAGCCGCAACCAAAAC 3317  
Qy 2428 TTCTCGGGAAGATGACGAAGAAAGATCTTTGATGACAGCAACCTCTCAATTTATCT 2487  
Db 3318 TTTAAAGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3377  
Qy 2488 TTGCTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2547  
Db 3378 ATCCCTGCTGAGATTCGTTGATGAAATATTCGCAAGATGAAATTAATTTGAGATTT 3437  
Qy 2548 ACTTATATCTATGCTTCCGATCTTATCCGCAATGATCCCAATGACACTACAGACTTTGA 2607  
Db 3438 TCTCTACCTATCATTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3497  
Qy 2608 ATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2667  
Db 3498 GTAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3557  
Qy 2668 CGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2727  
Db 3558 AGTCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3617

Qy 2728 GAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2787  
Db 3618 GAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3677  
Qy 2788 GAGGCTCTCTATGCTCA 2806  
Db 3678 TTCTTACTTCT 3696  
RESULT 15  
LOCUS AB1841 2793 bp DNA PAT 21-JAN-2000  
DEFINITION Sequence 15 from Patent WO9858953.  
ACCESSION AB1841  
VERSION AB1841.1 GI:6731871  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1. (bases 1 to 2793)  
AUTHORS Madsen A. and Birkelund S.  
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE  
JOURNAL Patent: WO 9858953-A 15 30-DEC-1998.  
MADSEN ANNA SOFIE (DK); BIRKELUND SVEND (DK)  
FEATURES  
source 1..2793  
location/Qualifiers  
BASE COUNT 803 a 681 c 567 g 742 t  
ORIGIN  
Query Match 23.0%; Score 646.4; DB 9; Length 2793;  
Best Local Similarity 54.6%; Pred. No. 2e-144;  
Matches 1539; Conservative 0; Mismatches 1221; Indels 60; Gaps 10;  
Qy 1 ATGAATCGCAATTTTCCTGTTAGTCTCTCTGACATTTGCAATGTTTACTAGTTGT 60  
Db 1 ATGAATATACCTTGCACAAACCTCTGATCTCTGACTCTTGACATCCCATCTATG 60  
Qy 61 TCCACTGTTTTCGTCGACATGCTGAAATATAGCCCTCTGATGATGATGATGATGATGATGAT 117  
Db 61 AGCATTTGCACTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120  
Qy 118 AGTACTAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 177  
Db 121 GCGGCTCTACATTTTACTCCAAATCTACAGCAGATGCAATGCAAGCACTATGCTTTA 180  
Qy 178 ACAGAGATATACCTCTGCAAAACCTTGGGATTCGCGAGCTTTTAAAGAGGTTGTTT 237  
Db 181 TCAGAAATATGCTATTAATTAACGATGCTGGGAAGCAAGCATTAACAGCGCTGCTT 240  
Qy 238 TCTACACTACGGAATCTTTAGCTTTGCGGTAAGGGGATCTACTGTTCTTTTAAAT 297  
Db 241 ACAGAAATCTAGGCTGATCTGCAATTTCTGGAAGGAAATCTATTTTCAATCAACAG 300  
Qy 298 ATTAAGTCTAGTGC---TGAAGGCGAGCACTTCTGTTCAACTGATTAATAAATCTGTGC 354  
Db 301 GTAGATGCGGCTTGAATGCGAGGAGCTGCGGCAAGCAACAGCTGATTAAGGCTTAA 360  
Qy 355 CTAAAGATTTTTCAGCTTATCTTTCTTAAGCGGCCCATCATCATGATTAACAAACCCC 414  
Db 361 TTACAGAGATTTTCTAACCTTTCTCTTCAATGACAGCTC-----CTGGAAGTACAGTTGCT 414  
Qy 415 TCAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 474  
Db 415 TCAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 474  
Qy 475 TTATTTAAACAAGATTAAGTGAAGAA-----AATGCGGAGGCAATTTTACCAAG 525  
Db 475 CTCTTTAGCCAAAGCTCTCCATGAAAGCTAATTAACATGCGGAGGAGGATCACCACAAA 534

QY	526	AAATCTTCTTTGAAAAACACGACGGATTCGATTTCTTTGAAGGAAATTAATTCAGCGCA	585
Db	535	ACTCTCTCTATTTCTGGGAAATACCTCTTCTATTAACCTTACTATGAATAGCGCAAAA--	591
QY	586	ACAGGGAAGAAAAAGTGGGCGCTATTTGTGCTAGTGGTACGTGTGATTAATACAAATATACG	645
Db	592	-----AAATTAGGTGGAGGATCTATTAACCTCTCGCGCGTCGCAAGATTTTCAGGAACACC	645
QY	646	GCCTCTACCCCTCTTCTCGAACAATATTGCTGAAGCTGCAGGTGAGCTATTAATAGCACA	705
Db	646	GGCCAGTTAGCTCTTTATGATATAATAAGGAGAAACTGGGGGGGGGGCTCTGGGCTTTGAA	705
QY	706	GGAACCTGTCAATTTACAGGGAATTAAGTCTCTGATTTTGGAAAATAGTGTACAGCG	765
Db	706	GCCAGCTCTCTGATTAACATAAATTAACCTCCCTTTCTCTCGAACAACGTCAACAGAT	765
QY	766	ACCGCAGGAAATGAGAGAGCTCTTCTGGAGA-----TGCCGATGTTACC	810
Db	766	GCCTGACGAGGAGCGCGGCTATTTGTGAAAAAACGAGAGACTCTACTCTTTACT	825
QY	811	ATATCTGGGAATCAGAGTGAATCTTCTCAGGAACCAAGCTGTAGCTATAGCGGGAGCC	870
Db	826	ATCTCTGGAAATTAAGTCTGACCTTCCCGCGAATCTTCACTGATCACTCAAGCGGAGCA	885
QY	871	ATTATGCTTAAGAGCTTACTGCTCTCCGGGGGGGGGGGTATCTCCTTTCTAAC	930
Db	886	ATCTGCTCCCATGCTGTAGATCTTTCTCGGTGGCGCC-----TACCCTATTTTCAAT	939
QY	931	AAATATATCCCAAGTATACACTCTGCAGATAATGTTGAGCCATTTCTATCTGGCAGCTGA	990
Db	940	AAATGAATGCGGGAACACAGCTGTGAGGCAAGGGCGGCGCTATTTGCAATTTGCGCAGCTGA	999
QY	991	GAGGTAGTCTTACACAGACAGAGGAGCAATTAACCTTCAATGGGAATGCCATTGTTCA	1050
Db	1000	TCTTTAGTCTCTCTGCAAAATCAAGAGACATCAGCTTCTTGGCAACATCTTAACCTCA	1059
QY	1051	ACTACACCAAACTCAAAAAAGAAATTTCTATTGACATAGATCTACTGCAAAATCACG	1110
Db	1060	ACCTCCCGCCCAATCGACACGGAATGCTACTACTCTGGGATCGTCGCAAAAATTTACG	1119
QY	1111	AAATTAGTGCATATCTGGGCACTAACCTCTTTTCTAGATCGGATTTACGTATATACG	1170
Db	1120	AACTTAAAGGCGCAAGCCCAAGCCCATATTAATCTTTCTATGATCGCAATTTGATCAACC	1179
QY	1171	GCCTCGGATTTCTACAGTACTTTTAAATCTCAATTAAGGCGTGAATGCAATTAAGTACAGAT	1230
Db	1180	ACAGAG--CTTCAGACGTTCTGACCATCAACCAACCGGATAGACATCCGCTTTAGAT	1236
QY	1231	TATAGTGGGTGAAATGTTTTTTTCTGTGAAAAGCTCTCGAAGATGAAGCAAAAGTTGCA	1290
Db	1237	TATTCAGGAAGATGATGATTTTCTGGGAAAAAGCTCTGCGAGATGAAGCAAAAGCTGCT	1296
QY	1291	GACAACTCACTTCTACGCTGGAAGAGCCTGTAACTCACTAACGTCAGAGAAATTTAGTACTT	1350
Db	1297	GATTAATCTTCAATCTATATTAATGAAGCAACCATTTGGCTTAGAGCTCTGGAACCTTAGCACTC	1356
QY	1351	AAACGTGTCTCACTCTCGATACGAAAGCTTTACTCAGACCGCGGGTCTCTGTTATTT	1410
Db	1357	AAAGGAATGTCAAGTAGTAGTGCATGTTTCAACAGACAGTGAAGGCTCTTACACTCTCTC	1416
QY	1411	ATGAGTACGGGCAACAGTTAAAGCAATATACAGAGAGGTACTTTAAAGGCTTTGCC	1470
Db	1417	ATGCAACAGGAACAAAGCTCAAGACAGATACTGAAGCTATCAGTCTTACCAAACTTGTCTC	1476
QY	1471	ATTCCTGTAGACTCTTTAGCGGAGGTAGAAGATTGTAATGCTGCTTCTGCAGCAGT	1530
Db	1477	GTTTATCTTTCTCGCTTAGAGGAAATTAAGAGTGTGTCATGAAGACAGCAGAGAGCCAC	1536
QY	1531	AAAAATGTACCCCTTATGTGTCGATCTTTCTTTTGGAAATCAACAGGAATGCTATTAGAA	1590
Db	1537	AAAACTTAATCTCACTCTCTCTTGTTTTCAGAGATGATGACGGCAATTTTATTAGAA	1596
QY	1591	AATCACAATTAAGAAAAATCAAGACTTTTATATTGTGACAGCTCTGCTCTGGGTACT	1650

Db	1537	AGCCATACGATTAACCAAGCCCTTCACGCAACCTTTGGTGGTATTTACATCTCTACTCTCT	1536
Qy	1651	GCACAACTACAGATGTCTTCACAGGGTTCCTACAGTAGTAGCAACTCTACGCACTATGGGTAT	1710
Db	1657	GCTAGCGATATTATTATTCGATGGCGCTTCACCTTCTCCAGTAGTACAAATCTCAGAACCTCAT	1716
Qy	1711	CAAGTACTTGGGGAATGACTTGGGTGATGATACCGCAAGCACTCCAAAGCTAAGACA	1770
Db	1717	TACGGGTATCAGGGA---CATTTGGGAAGCCACTTGGGCAAGACATCAACTCAGCAAAATCA	1773
Qy	1831	GTTCCTAATAGCCTTTGGGATCTTTTTCAGACATCCAAAGCGATTCAGGTCTCAFPAG	1890
Db	1834	GTTCGCCATTCATTATTTGGGATCCCTTCTTACGACATTCGCACTCTACAGCAATCATGACA	1893
Qy	1891	AGAAGTGGTTTGACTCTTTCTTCAGATCGAGGCTTCTGGGCTGGCGGAGTCCCAATTTTC	1956
Db	1894	TCTCAAGCGAATATATATCTATCAGCAACGAGCACTCTGGCATCAGAACTCGCAATTTTC	1955
Qy	1951	TTAGATAAAGATAAGAAAGGGGAAAAAGCAATACCGTATTAATCTGGTGGATATGCT	2010
Db	1954	TTCCATTAAGCATTAATACAGAACTAACCAAGCATTTCCGACATTAAGAAGTACGCGTATAT	2013
Qy	2011	ATCGAGGTGCAGCGCAACTTGTCTCGAAACTTAATTAGCTTTGCCCTTTGGCCACTC	2070
Db	2014	GTTCGAGGAAGTGCATGAAGATTTTCTCGAAAAATCTTCAAGTGTAGCTTTCTGCGCAGCTC	2073
Qy	2071	TTTGTAGCGATTAAGATTTTCTTAGTCGCTAAATAATCATACTGATACCTATCAGACAGCC	2130
Db	2074	TTCCGTAAAGATAAAGACGCTTTATAGTTGAATAAATACCTCATATACGATTTTACGTCG	2133
Qy	2131	TTCTATATCCACACATTTACAGAAATGATGGTTCATAGTTGTCTTATAGATAACTT	2190
Db	2134	CTATATCCATCGCAACATTCAGAGATTTCTTAGAGGAGCTTCCATGCCCTCATTTTGGAGTATC	2193
Qy	2191	CTCGGCTCTTGGATCTATTAACCCCTCGTTTATGAAGGGCAGCTCGCTATATAGCAAGTTC	2256
Db	2194	ACCGACATGCTGAAGAATATTTCTCTCATTTTAAATGCCAGCTAAGCTACAGCTACACT	2253
Qy	2251	AGTATGATGTGAAGACAAAGATTACTGCGTATCTGAGGTGAAGGTTCTTGGGGGAT	2310
Db	2254	AAAAATGATATGATCTCTGCTATCTTACTCTCATCTCGAAGCTCAAGGTTCTTGACCAAT	2313
Qy	2311	AATGCTTTTAAACAGATGTGGGAGCGCTCG---TCTCATTTCTTATCTGAAATACCTGAT	2367
Db	2314	AATTCGTGGGCTCTAGAGCTGGGAGATCTGTGGCTATATCTCTCTTAABABAGACATCG	2373
Qy	2368	TGTTTTGATACCTATGCTCATACATCAAACTGAATCTAGCTATATATAGTACGAGACAGC	2427
Db	2374	TTCTTCCAGGATATTTTCCCTCTTTAAAGTTCAGAGCAAGTCAACGCGCCACAAAC	2433
Qy	2428	TTTCTCGGAGAAAGGTACAGAAAGAAATCTTTGATGACAGCAACCTCTTCAATTATCT	2487
Db	2434	TTTTAAAGAGATGGCGCTGAAGCCCGTGCTTTGATGATGAGAGCACTAATGAATGCTCT	2493
Qy	2488	TTGCGCTAAGGGGAGATTTGGAAGTCTGATGTATAGACTTTTCTTATATATG	2547
Db	2494	ATCCCTGTGCGGCAATGGGTATGAAAAAATCTCGGAAGATGAAAAAATATTTTGCAGATT	2553
Qy	2548	ACTTTATCTATGTCTCTGATCTTATTCGCAATGATCCCAATGCACTACAGACTTGT	2607
Db	2554	TCTCTAGCAACAATTTGGTGTATGTATTCGTAATAAATCCCGCTTCGATCTTCTATATG	2613
Qy	2608	ATCAGCGGAGGCTCTTGGAAACTTATAGCAATATCTTGAACGACGACAGGCCCTTGAAGTG	2667
Db	2614	GTCAGTGAAGCTTCTTGACTTGGCTATGTAATAAACTCGCACGACAAAGCTTCTTACGA	2673
Qy	2668	CGTCAAGGCACTCATCAGGCTTCTCTCTATGTTTGAAGTGCCTCGGCAAGTTGTCTTT	2727

Db 2674 AGTGTGGAAGCCATCTGACTCTCTCCCTCATGTAGAACTCTCTGGGGAAGCTGCTTAT 2733  
QY 2728 GAAGTTCGTGATCCCTCAGGATTTATATGTAGATCTTGGGGTAAAGTCCAAATTTCTAG 2787  
|| ||||| || || || ||||| ||| | | |||||  
Db 2734 GAGCTTCTGGCTCAGCACACATCTACATGTAGATTGTGGCTAAGATACTCATTTCTAG 2793

Search completed: October 2, 2001, 00:43:56  
Job time: 43309 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 03:56:46 ; Search time 800.41 Seconds  
(without alignments)  
2208.297 Million cell updates/sec

Title: US-09-446-677B-3

Perfect score: 2815  
Sequence: 1 ATGAATGCGCAATTTCCTG.....CTCAGTCCTCAGAAATTCCTG 2815

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.Geneseq\_0601:\*

- 1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT:\*
- 2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT:\*
- 3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT:\*
- 4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT:\*
- 5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT:\*
- 6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT:\*
- 7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT:\*
- 8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT:\*
- 9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT:\*
- 10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT:\*
- 11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT:\*
- 12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT:\*
- 13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT:\*
- 14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT:\*
- 15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT:\*
- 16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT:\*
- 17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT:\*
- 18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT:\*
- 19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:\*
- 20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT:\*
- 21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:\*
- 22: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2815	100.0	2815	20 AAX06817	Chlamydia pneumoni
2	2803	99.6	273254	21 AAC81914	Chlamydia pneumoni
3	2771.8	98.5	2950	21 AAX30849	Chlamydia antigen
4	2742	97.4	3000	20 AAX06828	Chlamydia pneumoni
5	2683.4	95.3	2784	21 AAX30850	Chlamydia antigen
6	672.2	23.9	2957	21 AAX61509	DNA encoding the C
7	670.6	23.8	2757	20 AAX06821	Chlamydia pneumoni
8	649.4	23.1	273254	21 AAC81914	Chlamydia pneumoni
9	649.4	23.1	1230025	20 AAX91990	Nucleotide sequence
10	647.8	23.0	3000	21 AAX30853	Chlamydia antigen
11	646.4	23.0	2793	20 AAX06823	Chlamydia pneumoni

12	615.4	21.9	2790	21 AAX30854	Chlamydia antigen
13	536.2	19.0	2950	21 AAX30851	Chlamydia antigen
14	535	19.0	2787	20 AAX06820	Chlamydia pneumoni
15	530.2	18.8	3000	21 AAX27021	Chlamydia pneumoni
16	527.4	18.7	3050	21 AAX61508	DNA encoding the C
17	521.8	18.5	2787	20 AAX06822	Chlamydia pneumoni
18	519	18.4	2784	21 AAX30852	Chlamydia antigen
19	518.2	18.4	2838	20 AAX06827	Chlamydia pneumoni
20	456	16.2	2950	21 AAX30847	Chlamydia antigen
21	456	16.2	3200	20 AAX06816	Chlamydia pneumoni
22	446	15.8	3050	21 AAX48838	Chlamydia antigen
23	427	15.2	2784	21 AAX30848	Chlamydia pneumoni
24	422.2	15.0	2808	21 AAX48839	Chlamydia pneumoni
25	309.8	11.0	2750	20 AAX28690	C. pneumoniae CPN1
26	309.4	11.0	2526	20 AAX06819	Chlamydia pneumoni
27	307.8	10.9	2526	21 AAX28691	C. pneumoniae CPN1
28	233.2	8.3	2950	21 AAX50035	DNA encoding Chlam
29	231.6	8.2	3052	20 AAX06818	Chlamydia pneumoni
30	214.2	7.6	1830	20 AAX06830	Chlamydia pneumoni
31	171.6	6.1	3150	21 AAX27342	Chlamydia POMP91B
32	170.6	6.1	840	20 AAX06824	Chlamydia pneumoni
33	125.8	4.5	1200	20 AAX06829	Chlamydia pneumoni
34	114.8	4.1	3021	21 AAX64758	C. trachomatis pmp
35	114.4	4.1	2949	21 AAX64752	C. trachomatis pmp
36	113.2	4.0	1444	20 AAX60547	C. trachomatis HmV
37	113.2	4.0	4435	20 AAX60539	C. trachomatis Icy
38	109	3.9	3354	20 AAX60556	Seq ID No: 23 of W
39	108.2	3.8	3324	20 AAX60557	Seq ID No: 24 of W
40	95	3.4	936	22 AAF58252	Oligonucleotide D1
41	95	3.4	936	22 AAF58254	Oligonucleotide D1
42	95	3.4	936	22 AAF58255	Oligonucleotide D1
43	95	3.4	936	22 AAF58259	Oligonucleotide D1
44	95	3.4	936	22 AAF58262	Oligonucleotide D2
45	95	3.4	938	22 AAF58255	Oligonucleotide D1

#### ALIGNMENTS

RESULT 1	
AAX06817	standard; DNA: 2815 BP.
ID	AAX06817
XX	
AC	AAX06817:
XX	
DT	26-APR-1999 (first entry)
XX	
DE	Chlamydia pneumoniae surface exposed protein Omps DNA.
XX	
KW	Omps: outer membrane protein 5; surface exposed protein; antigen;
KW	Infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
OS	Chlamydia pneumoniae.
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
XX	1..2787
XX	/*tag- a
XX	
XX	WO9858953-A2.
XX	
PD	30-DEC-1998.
XX	
PF	19-JUN-1998; 98WO-DK00266.
XX	
PR	23-JUN-1997; 97DK-0000744.
XX	
PA	(BIRK/) BIRKELUND S.
XX	(CHRI/) CHRISTIANSEN G.
XX	
PI	Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX	Mygind P;
DR	WPI; 1999-105610/09.



DR P-PSDB; AAW88418.

XX Species-specific test for identifying mammals infected with  
PT Chlamydia pneumoniae - comprises detecting antibodies specific for  
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding  
PS these proteins

XX Claim 6; Page 42-43; 115pp; English.

CC This DNA sequence codes for the novel 97.2 kDa surface exposed  
CC protein Omp5 (see AAW88418) of the human respiratory pathogen  
CC Chlamydia pneumoniae. By generating antibodies against C.  
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)  
CC was obtained which reacted with outer membrane proteins. The  
CC antibody was used to identify the genes (see AAW06816-27) encoding  
CC Omp4-Omp5 proteins (see AAW88417-28) in an expression library of  
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:  
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp5,7,8,9 and 15 in  
CC the other, and encode polypeptides of about 89.6-100.3 kDa and  
CC about 56.1 kDa. The invention provides a new species specific test  
CC for identifying mammals (including humans) infected with Chlamydia  
CC pneumoniae. The test comprises detecting antibodies specific for  
CC Omp4-Omp5 or detecting nucleic acid fragments encoding these outer  
CC membrane proteins, especially by PCR. The proteins are also used  
CC in the diagnosis of C. pneumoniae infection in mammals. The  
CC nucleic acids and proteins can also be used in the immunization of  
CC mammals, the nucleic acids being particularly useful as DNA  
CC vaccines for effecting in vivo expression of antigens. The  
CC vaccines may also prevent atherosclerosis and bronchial asthma,  
CC which are possibly associated with C. pneumoniae.

XX Sequence 2815 BP; 790 A; 575 C; 626 G; 824 T; 0 other;

Query Match 100.0%; Score 2815; DB 20; Length 2815;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2815; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAATCGCAATTTCTGTTAGTCTCTCTCGACATTGGCATGTTTACTAGTGT 60  
DB 1 atgaatcgcgaatttcttcttgtagtctctctgcacatgtagtcttcttactagttg 60  
QY 61 TCCACTGTTTTGCTGCACTGCTGAATAATATAGGCCCTCTGATAGCTTTGACGAGT 120  
DB 61 tccactgttttctgcaactgctgaataatagagccctctgtagcttgcagcgaagt 120  
QY 121 ACTAACACAGGACACTTACTCTTAAATAATACGACTACTGGAATAGCTATCTACACA 180  
DB 121 actaacacaggcacttactcttctctctctctctctctctctctctctctctctct 180  
QY 181 GGAAGATATTAATCTGCAAAACCTTGGGATTTCGACACTTAAAGAGGTTGTTTCT 240  
DB 181 ggaagatataactctcgaacaaaccttgggatttcgacactttaaagaggtgttttct 240  
QY 241 GACACTAGGGAATCTTAAAGCTTTGCCGTAAGGGTACTCACTTTCTTTTAAATATT 300  
DB 241 gacactagggaaatctttaaagcttttgccgttaagggactcaacttttctttaaataat 300  
QY 301 AACTAGTACTGCTGAAGCGAGACATTCTGTATACACTATATAAATCTGTGCTTACA 360  
DB 301 aactagtagtgcgaagcgagacattctgttatacaactatataaatactgtgcgtaca 360  
QY 361 GGAATTTTCGAGTCTACTTTCTTAGCGGCCCATCATCGGTAATCACACCCCTCGAGA 420  
DB 361 ggaattttcgagtctacttttcttagcgggcccatcatcggtatacaaacccctcga 420  
QY 421 AAAGGTGCACTTAAATGTGAGGGGATCTTACATTTGATAAACAATGGAATTTTATTT 480  
DB 421 aaaggtgcactttaaagtgtgaggggatcttaccttatacaaatggaattttattt 480  
QY 481 AAACAAGATTACTGTAAGGAAATGGCGAGCATTTCTACCAAGAACTCTTCTTAA 540  
DB 481 aaacaagattactgtaggaataatggcgagcatttctacccaagaatcttcttctgaaa 540

QY 541 AACAGCAGGGATCGATTCTTTGAAGGGAATTAATCGACGCACAGGAAAAAAGT 600  
DB 541 aacagcagggatcgatttcttgaagggaatttaatactcgacgcacaggaaaaaagtc 600  
QY 601 GGGCTATTGTGCTACTGCTGATGATTTTCAAAATTAATACGGCTCTTACCTCTTC 660  
DB 601 gggctattgtgctactgctgattgatttcaaaaataatagctcttctctctctc 660  
QY 661 TCGAACAAATATTTGCTGAAGCGCAGTGGAAGCTTAATAGCACAGGAACCTGACAA 720  
DB 661 tcgaacaataatctgtagaagctgcaggtggaagctataaataagcaagaaactgcata 720  
QY 721 ACAGGGAATACGTTCTTGTATTTTGAATAATAGTGTGACAGCGACGACGAAATAGA 780  
DB 721 acagggaatacgttctgttatttcttgcgaataatagtgtagcagcgaccgagaata 780  
QY 781 GGAAGCTTTCTGAGATGCCGATGTTACCAATCTGGGAATCAGAGTGTAACTTTCTCA 840  
DB 781 ggaagcttcttgagatgccgatgtttaccatcttggaatcagagtgtaacttctca 840  
QY 841 GGAACCAACCTGAGCTAATGGCGAGCCATTATGCTAAGAACTTACACGCTTCC 900  
DB 841 ggaaccaacctgagctaatggcgagccattatgcttaagaaacttacacgcttcc 900  
QY 901 GGGGGGGGGGGGATCTCTCTTTCTAACAAATATATCCAGGTACACTGACGTAAT 960  
DB 901 ggggggggggggatctctcttcttcttcttcaacaataatccaggtaacctgcagaa 960  
QY 961 GGTGAGCCATTTCTTACTGACGAGCTGAGAGTGTACTTTTACGACGAAGCGGAG 1020  
DB 961 ggtgagccatttcttactgacgagctgagagtgtacttttcaagaagaagcgaggac 1020  
QY 1021 ATTACCTTCAATGGGAATGCAATGTCGAACATATATCTGGAATATCTTCTTCT 1080  
DB 1021 attaccttcaatgggaatgcaatgtcgaacatattctggaatattcttcttcttct 1080  
QY 1081 ATTGACATAGGATCTACTGCAAAAGATCAGCAATTTACGTGCAATATCTGGCATATG 1140  
DB 1081 attgacataggatctactgcaaaaagatcagcaatttacgtgcaatattctggcatatg 1140  
QY 1141 TTTTCTTCAAGATCCGATTTACTGCTAATAGCGCTCGGATTTCTACAGTATTTAA 1200  
DB 1141 ttttcttcaagatccgatttactgctaatagcgctcggtttctacagtattttaaactc 1200  
QY 1201 AATAAGGCTGATGAGGTAATAGTACAGATTTATAGTGGTGGATTTCTTCTGCGAA 1260  
DB 1201 aataaggctgatgaggttaattagtacagatttatagtggtggatcttcttctgctgaa 1260  
QY 1261 AAGCTCTCTGAGAGTGAAGCAAAAGTTGACAGAACCTCACTTCTACGCTGAAGCAG 1320  
DB 1261 aagctctctgagagtgaagcaaaaagttgacagaaacctcaacttctacgctgaagcag 1320  
QY 1321 GTAACTTAATCGCAGGAATTTAGTACTTAAACGTGTCTACCTCGATACGAAAGGC 1380  
DB 1321 gtaactttaatcgcggaatttagtactttaaagctgtgtctctcgaatacgaagc 1380  
QY 1381 TTTACTAGACCGGGGTTCTCTGTATATGAGATCGGGCAACGTTAAAGCAAGT 1440  
DB 1381 tttaactagaccggggttctctgtatattagatcgaggacggttaaaagcaagt 1440  
QY 1441 ACAAGAGAGTCACTTTTAAAGCTTTTCCATTCTGTAGACTCTTTAGCGGAGGTAAG 1500  
DB 1441 acagagaggtcacttttaagcttttccattctgttagactctttagcgaggtgaag 1500  
QY 1501 AAAGTGTATTTGCTGTTCTGACGACAGTAAATATATACCTTACTGATCCGATCTT 1560  
DB 1501 aaagtgtatttgcgttctgacgacagttaaatattaccttactgattccgattctt 1560  
QY 1561 CTTTGGATTACCAAGGAATGCTTATGAATAATCACACATTAGGAATAAATCAAGACTT 1620  
DB 1561 ctttggattaccaaggaatgcttattgaataatcacacatttaggaataaatacaagactt 1620

QY 1621 TCATTGTGACGCTCTGCTGTGGTACTGCAACAACATGATGTTCCAGCGCTTCT 1680  
 |||||||  
 Db 1621 tcaattgtgacgctctctgtctgtggtactgcacaactacagatgtctcagcggttcc 1680  
 |||||||  
 QY 1681 ACAGTAGCAACTCTCTACGCACTATGGGTATCAAGGTACTGGGGATGACTTGGGTTGAT 1740  
 |||||||  
 Db 1681 acagtagcaactctctacgactatggtatcaagtagtacttgggaatgacttgggttgc 1740  
 |||||||  
 QY 1741 GATACCGCAAGCACTCCAAAGACTAGACAGCGCATATTACTTGGACCATACAGGCTAC 1800  
 |||||||  
 Db 1741 gataccgcaagcactcccaagactaagacagcgacattagcttggaccaatacaggtac 1800  
 |||||||  
 QY 1801 CTTCGCAATCTGAGCGTCAAGGACCTTGTCTTAATAGCCCTTGGGATGCTTTTCA 1860  
 |||||||  
 Db 1801 ctccgcaatctctgagcgctcaagaccttagtctcctaagcttggggatgacttcttca 1860  
 |||||||  
 QY 1861 GACATCCAAAGCGATCAAGGTGTCATAGAGAGAAGTCTTGTGACTCTTCTTTCAGATCGA 1920  
 |||||||  
 Db 1861 gacatccaaagcgatcaaggtgtcatagagaagagtgcttgcactcttcttgatcagatcga 1920  
 |||||||  
 QY 1921 GCGTCTGGGCTCGGGAGTCCCAATTTCTTAGATTAAGATAAGAAAGGAAAAAGC 1980  
 |||||||  
 Db 1921 ggcctcgggctcgggagtcgccaatttcttagataaagaaggggaaaaagc 1980  
 |||||||  
 QY 1981 AATPACCGTATTAATCTGATGATATGCTATGAGGAGTGCACGCGCAACTGTTCTGAA 2040  
 |||||||  
 Db 1981 aatpaccgtatataatctgatatgctatgagggagtcacgcaactgttctgaa 2040  
 |||||||  
 QY 2041 AACTTAATTAAGCTTTTGCTTTTGCACACTCTTTTGATAGCATTAAGATTTCTTAGTCGT 2100  
 |||||||  
 Db 2041 aacttaattagcttgccttcttgccaactcttcttgatagataaagttcttcttagtcgtc 2100  
 |||||||  
 QY 2101 AAAAATCATCTACTATCTATCTATGCAAGAGCTTCTATATCCACACATTTACAGATGATG 2160  
 |||||||  
 Db 2101 aaaaatcatctactatctactatgacagagccttctatataccaacattacaagaatgagt 2160  
 |||||||  
 QY 2161 GGGTTATAGTGTCTCTTAGATAAATCTCGGCTCTGAGTCAATAACCCCTGCT 2220  
 |||||||  
 Db 2161 gggttatagtgctctcttagataaactcctcgcccttggagtcataaaccctgct 2220  
 |||||||  
 QY 2221 TTAGAAGGAGCGCTCGCTATATGCCACGTCACTATGATGATGTAAGACAAATTAATCTGCG 2280  
 |||||||  
 Db 2221 ttagaaggagcgctcgctcttagccacgtcagtaagatctgaaagacaaagtactacgcg 2280  
 |||||||  
 QY 2281 TATCTGAGGTGAAGGCTCTTGGGGATTAATGCTTTTAACATGATGTTGGAGCTTCT 2340  
 |||||||  
 Db 2281 tatctgaggtgaaaggtcttggggaaatgcttttaacatgattgttggagcttct 2340  
 |||||||  
 QY 2341 TCTCATTTCTTATCTGATTAATCTCGCATTTGTTTGATACCTATGCTCCATCATCAAACTG 2400  
 |||||||  
 Db 2341 tctcatcttctatctcctaatactcgtcattgtttgatacctatgctccatcatacctaactg 2400  
 |||||||  
 QY 2401 AATCTGACCTATATAGTGCAGAGAGCTTCTCGAGAAAGTACAGAAAGAAATCTTTT 2460  
 |||||||  
 Db 2401 aatctgacctatatagtcagcagcgtctctcgagaaagatcagaagaagatcttct 2460  
 |||||||  
 QY 2461 GATGACAGCACTCTTCAATTTATCTTGGCTTAAGGGGTGAAGTTTGAAAGTTCTCT 2520  
 |||||||  
 Db 2461 gatgacagcaactcttcaattatcttgcctataggggtgagtttgaaagttctctc 2520  
 |||||||  
 QY 2521 GATTGTAAAGACTTTTCTTATGATCTGACTTTTCCATGTTCTGATCTTATCCGAAT 2580  
 |||||||  
 Db 2521 gatgttaaagacttttcttattgactgacttatacctatgcttccatctcgaat 2580  
 |||||||  
 QY 2581 GATCCCAAAATGCACTACAGCACTTGTATATCAGCGAGACCTCTTGGGAAACTTATGCCAAT 2640  
 |||||||  
 Db 2581 gatcccaaaatgcaactacagactgtfaatcagcgagcctcttgggaaacttattgcaat 2640  
 |||||||  
 QY 2641 AACTTAGACAGCAAGGCTTGGCAAGTGCCTGCGAGCAAGTACATACGCTTCTCTCTATG 2700  
 |||||||  
 Db 2641 aacttagcagcaagcgcttgcgaagtgcgagcgagtcactacagccttctctcatatg 2700  
 |||||||  
 QY 2701 TTGGAAGTGTGCGCAAGTTTGTCTTGAAGTCTGTGATCCCTCAGCGAATTTAATGTA 2760

Db 2701 ttggaagtgtcgccaggttgcctttagaagttcgtgacctccacagattataatgta 2760  
 |||||||  
 QY 2761 GATCTTGGGGGTAAAGTTCCAAATTTAGAGCGGTCTCTCAATGTCACAAATTCG 2815  
 |||||||  
 Db 2761 gatcttgggggtaaagtccaatttagagcgctctcctcatgctcagaatctcg 2815

## RESULT 2

AAC81914  
 ID AAC81914 standard; DNA: 273254 BP.

AC AAC81914;

DT 27-FEB-2001 (first entry)

DE Chlamydia pneumoniae genome DNA.

KW Genome; diagnosis; vaccine; ds.

OS Chlamydia pneumoniae.

PN WC200027994-A2.

PD 18-MAY-2000.

PF 12-NOV-1999; 99MO-US26923.

PR 12-NOV-1998; 98US-0108279.

PR 08-APR-1999; 99US-0128606.

PA (REGC ) UNIV CALIFORNIA.

PI Stephens R, Mitchell W, Kalman S, Davis R;

DR WPI: 2000-376516/32.

PT Isolated nucleic acid for use in diagnostic and analytical methods

PS encodes genomic sequence of Chlamydia pneumoniae -

Claim 2: Page 128-320; 320pp; English.

CC This invention describes a novel nucleic acid (NI) encoding a Chlamydia  
 CC pneumoniae protein (PI), given in the specification. The isolated nucleic  
 CC acid is useful for diagnostic and analytical methods, such as,  
 CC hybridization-based assays or amplification-based assays. The protein may  
 CC be used for diagnostic purposes, for their enzymatic or structural  
 CC activity, or as a vaccine. The invention also describes (1) a probe  
 CC comprising a hybridizing fragment of NI; (2) an isolated nucleic acid  
 CC (N2) that hybridizes under stringent conditions to NI; (3) an expression  
 CC cassette comprising NI under the transcriptional regulation of a  
 CC transcriptional initiation region functional in an expression host, and a  
 CC cassette of (3) as part of an extrachromosomal element or integrated into  
 CC the genome of a host cell as a result of induction of the expression  
 CC cassette into the host cell, and the cellular progeny of the host cell;  
 CC (5) a method for producing a PI comprising growing a cell of (4) where  
 CC the protein is expressed and isolating the protein free of other  
 CC proteins; (6) a purified polypeptide composition comprising at least 50  
 CC weight % of PI; and (7) a monoclonal antibody binding specifically to the  
 CC peptide of (6).  
 CC  
 XX Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;  
 SQ

Query Match 99.6%; Score 2803; DB 21; Length 273254;  
 Best local Similarity 100.0%; Pred. No. 0;  
 Matches 2814; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ATGAAATCGCAATTTCTGTTAGTCTCTTCGACATTTGCAGTGTACTAGTTGT 60  
 |||||||  
 Db 98194 atgaatcgcaatttcttctgttagtctctcttcgacatgtcatgtttactagtgtg 98253



Dh	100413	ttggaaggcgacgtcgcgtctatagccaaagtcgaatgatctgaaagcaaatgactcgg	100412
Oy	2281	TATCCTGAGGTGAAGAGTTCTTGGGGGAATATAGCTTTTAAATGATGTTGGAGCTTCT	2340
Dh	100473	tatcctgagtgaaaggtctcttggggaaataatgctcttaacatgatgttggagctctt	100532
Oy	2341	TCCTATTCTTATCCTGAAATACCTGCATTGTTTGTATACCTATGCTCCATACATCAAACTG	2400
Dh	100533	ttcattcttctccctgaataccctgcattgttcttgataccatgctccataccataaacy	100592
Oy	2401	AATCTGACCTTATACCTGCAGAGAGCTTTCGGAGAAAGTACAGAGAAATATCTTTT	2460
Dh	100593	aacttgacctataatagtcacggagaaagcttcggagaagaagtacagaagaagatcttt	100652
Oy	2461	GATGACAGCAACCTCTTCAATTATATCTTGGCCATATAGGGGTGAAGTTTGAAGATTCTCT	2520
Dh	100653	gatgacagcaacctctccaattctatctcttgctataggggtgaagtttggaaagtctct	100712
Oy	2521	GATTGTATATGACTTTTCTTATGATCTACATTATACCTATATGTTCTGATCTTATCCGAAT	2580
Dh	100713	gattgtaatgacttctcttctatgacatcttaccatctctatcttccatcttaccgaat	100772
Oy	2581	GATCCCAATACACTATACAGACCTGTATATAGCGGACCCCTTGGGAAATTTATGCAAT	2640
Dh	100773	gattcccaaatgactactaagaacactgtatatacagcgagccctctgggaaactatgccaat	100832
Oy	2641	AACATTAGCAGACAGGCGCTTGCAAGTGCATGACAGGACAGTCACTACGCGCTTCTCTCATG	2700
Dh	100833	aacttagagcaagaagcgctcttgcaagtcgfygcagcgactcaatacagctctctctatg	100892
Oy	2701	TTTGAAGTGTCTGGCCAGATTGTCTTTGAAGTTCGTGATGATCTCAGGATTTATTAATGA	2760
Dh	100893	tttgaagtgtcgcgcagttctgtctcttgaagttctgataccccaagattataatgta	100952
Oy	2761	GATCTTGGGGTATGTTCCAAATCTATAGACGCTCTCATGTCATCCGAATTTCTG	2815
Dh	100953	gattcttggggtaagttccaattcttaagagcgtctctccaagtctcagaactcttg	101007

Result	3
AAA30849	standard; DNA; 2950 BP.
AAA30849;	
29-AUG-2000	(first entry)
Chlamydia antigen CPN100635	full length coding sequence.
Chlamydia antigen; diagnosis; infection; community acquired pneumonia;	
therapy; upper respiratory tract disease; bronchitis; sinusitis;	
asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;	
Chlamydia pneumoniae.	
Key	Location/Qualifiers
CDS	101..2887
/*tag- a	/product= Chlamydia antigen CPN100635
101..229	
/*tag- b	
230..2884	
/*tag- c	
/note= "mature CPN100635"	
08-JUN-2000.	
01-DEC-1999;	99MO-CA01147.
01-DEC-1998;	9805-0110339.

PR 01-DEC-1998; 98US-0110340.  
PR 01-DEC-1998; 98US-0110427.  
PR 01-DEC-1998; 98US-0110428.  
PR 01-DEC-1998; 98US-0110438.  
XX  
PA (CONN-) CONNAUGHT LAB LTD.  
XX  
XX  
PI Murlin AD, Oomen RP, Wang J;  
XX  
XX WPI: 2000-412339/35.  
DR P-PSDB: AAY90237, AAY90238.  
XX  
XX  
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for  
PI preventing, diagnosing and treating diseases such as community acquired  
PI pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset  
PI asthma -  
XX  
PS Claim 2, Fig 3, 174pp; English.

CC This sequence encodes a Chlamydia antigen of the invention, designated  
CC CPN100635. The nucleic acids (and their complementary sequences) may be  
CC used as diagnostic agents for detecting the presence of nucleic acids  
CC encoding Chlamydia antigens in samples according to standard methods,  
CC and therefore, for diagnosing Chlamydia infections. For example, they may  
CC be used as primers and probes for diagnostic polymerase chain reaction  
CC (PCR) assays. Antisense sequences may be used to down regulate  
CC expression of the proteins and may be used to treat infections. The  
CC nucleic acids may also be used to produce the protein antigens they  
CC encode according to standard recombinant DNA methodologies. The  
CC proteins may then be used as antigens for the production of antibodies  
CC (i.e. as vaccines) for preventing infection by Chlamydia. The  
CC antibodies may also be used as diagnostic reagents for detecting  
CC infections. Chlamydia is a pathogen implicated in the development of  
CC (for example) community acquired pneumonia, upper respiratory tract  
CC disease (especially bronchitis and sinusitis, asthmatic bronchitis,  
CC adult-onset asthma and acute exacerbations of asthma in adults.  
XX  
SQ Sequence 2950 BP; 836 A; 593 C; 652 G; 869 T; 0 other;

Query Match	98.5%	Score 2771.8	DB 21	Length 2950
Best Local Similarity	99.0%	Pred. No. 0		
Matches 2788	Conservative	0	Mismatches 27	Indels 0
				Gaps 0
QY 1	ATGAATGCGCAATTTTCCGTGTAGTGCCTCTTGACATTTGGCATGTTTACTAGTTGT	60		
DB 101	atgaatcgcaattttccgtgtagtgcctcttcgcgaattgcatglttaactagtgt	160		
QY 61	TTCAGTGTTTTGCCTGCAACTGCTGAAAATATAGCCCTCTGATAGCTTTAGCGAGCT	120		
DB 161	tcacatgtttttgcgcgaactctgtaaaataagagcccttcgataagcttttcagcgaagt	220		
QY 121	ACTACACAGCGACCTATACCTCTTAAAAATATGACTATACGAAATGACTATCTCTGACA	180		
DB 221	actaacacagcgacactatactctctaaaaataagactactcgtaatgactatactctgaca	280		
QY 181	GGAGATATTAACCTCTGCCAAAACCTTGGGGATTCGGACGCTTTAAGCAAGGTTGTTTTCT	240		
DB 281	ggaagataaacctctgcaaaaccttgggagtttcgcgaacttaacgaaggttcttttct	340		
QY 241	GACACTAGGGAATCTTTAAGCTTTGCCGTAGGGGTACTGACTTTCTTTTAAATAT	300		
DB 341	gacactaaggaaatctctaagctcttcgcglaaggggtactcaacttcttttaaatatc	400		
QY 301	AACTAGTGTGCGAAGGGCGACAGCACTTCTCTTACAACTGATAAAAATCTGTGCTAAC	360		
DB 401	aagctagtgctgaagcgcgaccccttcttcgttaacaactgataaaaactctgcgttaaca	460		
QY 361	GGATTTTGGAGCTTACTTTCTTACCGGCCCATCATGGGTATATCAACAACCCCGCAGGA	420		
DB 461	ggattttcgagcttaacttctcttagcggcccatcatcgttaatacaacccccctcaagga	520		
QY 421	AAAGGTGACGTTAAATGTGGAGGGGATCTTACATTTGATATACAAATGGAATTTTATATT	480		

```
|||||
Db 521 aaagtgcaatgaatgaggggacatctacatcttgatacaaatgaaactatttattc 580
QY 481 AAACAAGATTACTGTGAGAAATGCGGAGCCCTTTCTACAAAGATCTTTCTTGA 540
Db 581 aaacaagaattactctgagaaatcgcgagccattctcacaagaattcttcttgaa 640
QY 541 AACAGCAGGGATGATTTCTTTTGAAGGAAATTAATCGAGCCGACAGGAAAGAGT 600
Db 641 aacagcagggatcattcttcttgaaaggaataatcgcgcgaacaggaagaaagyt 700
QY 601 GGGGCTATTTGTGCTACTGTGCTACTGTAGATTAATAAATATACGGCTCTACCTTTC 660
Db 701 ggggctatttgctgactgctgactgactgataataataagctccatccctctc 760
QY 661 TCGACAATATTTGCTGAAGCTGCGAGGTGAGCCTTAATATGACACAGAACTGTACATT 720
Db 761 tcgaacaatatctgcaagctgcagctgagctataatagcacaagaaactgtaaat 820
QY 721 ACAGGAATAGCTCTCTGATTTTCTGAAAATAGTGTGACAGCGACGAGAAATGGA 780
Db 821 aacaggaaatacgtctctgtatcttctcgaataatagtgtagaagcagcgagaaatgga 880
QY 781 GGAAGCTTTTCTGAGATGCGGATGTTACCATATCTGGGAATCAGACTGTAACTTTCTCA 840
Db 881 ggaagcttctcggagatgcgcgactgatacatctggaatcagagtgtaactctca 940
QY 841 GGAACCAACAGCTGTAGTAAATGGGAGCCATTATGCTAAAGAACTTACACTGCTTCC 900
Db 941 ggaaccaaagctgaagctaaagcgagccaatctgataagaaagcttaacatggtctcc 1000
QY 901 GGGGGGGGGGGGGGATCTCTCTTTCTAACAAATATAGTCCAAAGGTACACTGCAGTAAT 960
Db 1001 ggggggggggggggaaatccctcttctcaaatatagttccaaagttaccacgtcgagtaat 1060
QY 961 GGTGAGCCATTTCTATPACTGAGCTGAGAGTGTAGTCTTTGACGAAAGCAGGGAC 1020
Db 1061 ggtgagccattctcatcactgcagcgagagtgtagctcttccagcgagagcgagac 1120
QY 1021 ATATCTTCAATGGAGATGCAATGTTGCAACTACACCAAACTTACAAAAGAAATCT 1080
Db 1121 catcaacttaaggaatgcattgtgcactacacacaacaatacaaaaagaattctc 1180
QY 1081 ATTCACATAGCATCTACTGCAAAAGATCAGAAATTTACGTGCAATATCTGGCATAGCATC 1140
Db 1181 attgacataagatctactgcaaaagatcacgaattacgtgcaatactggtgcataagatc 1240
QY 1141 TTTTCTACGATCGATTACTGCTAATACGGCTGCGATTCTACAGTACTTTAAATCTC 1200
Db 1241 ttttctacgactcgaattactgctaatacgcgtgcgattcttaagaatactttaaatctc 1300
QY 1201 AATAAGCTGATGAGTAATATAGTATATAGTATATAGTGGTGGATTTTCTTGGTGAA 1260
Db 1301 aataagctgataagtaataatagtaacagattatagtggtgcgattgttcttctggtgaa 1360
QY 1261 AAGCTCTCTGAAGTAGAAGAAAGTTGACAGCAACCTCTCTACGTGAGAGCACT 1320
Db 1361 aagctctctgaagatgagaaagtgcagacaactctactacgtgaaagcagct 1420
QY 1321 GTAACTCTAATGAGAAATTTAGTAAAGGTGTGTCTCTGATAGGAAAGGC 1380
Db 1421 gtaacttaactcgcagaaatttagtacttaaacgtgtgtcaactctcgaatacgaagc 1480
QY 1381 TTTTACAGACCGGGGTCTCTGTTATATGAGATGCGGACAAAGGTTAAAAAGCAAGT 1440
Db 1481 tttaactaagccgcgggtctcgtctgtaataagtagcgggcaaaagcttlaaaagcaagt 1540
QY 1441 ACAGAGAGGATCTTAAACAGGCTTTCCATTCTCTTAACTCTTTAGGCGAGGAGT 1500
Db 1541 acagaggaagtgtaacttaaaaggtcttccattccgttgaactcttaaggcgaggtgaa 1600
QY 1501 AAGGTGTAAATGCTGCTTCTGACGAAGTAAAAATGTAGCCCTTATGTGCTCGATTCTT 1560
|||||

|||||
Db 1601 aaagtgtaatctgctctctcgcagcaagtaaaaatgtaagcccttagtggtccgattctc 1660
QY 1561 CTTTTGATACCAAGGAATGCTTATGAATAATCACACTTAGGAAAAACATCAAGACTTT 1620
Db 1661 ctcttgataaccaaggaagctctatgaaatacaagacttaggaaataactcaagactc 1720
QY 1621 TCATTTGCGAGCTCTCTGCTGCTGCTACTGCAACAACATACAGATGTTCCAGCGGTTCT 1680
Db 1721 tcaatttgagagctctctcgtctcgtgtaactgcgaacaactacagatgtctccagcgcttc 1780
QY 1681 ACAGTACCAACTCTCTACGCACTATGGTATACAGAGTACTTTGGGAAATGACTTGGTTGAT 1740
Db 1781 acagtagcaactctcctacgcactagtggtatcaaggtactctggggaaatgacttggtgat 1840
QY 1741 GATACCGCAAGCACTCCAAAGACTAAGACAGCGACATTTAGCTTGAGCAATPACAGGCTAC 1800
Db 1841 gatacgcaagcaactccaaagaaactaagaacgagacattagcttggaaccaatacaggtctac 1900
QY 1801 CTTTCGAATCCTGAGCGTCAAGGACCTTTAGTTCCTAATAGCCTTTGGGATCTTTTCA 1860
Db 1901 ctccgaatccctgagcgctcaagagcccttagttccataagccttggtggactcttca 1960
QY 1861 GACATCCAAAGCAATCAAGGTGATATGAGAGAAAGTCTTGAATCTTTGTCAGATCGA 1920
Db 1961 gaactcaagcgaltcaagtgltcatagagaagtgcttgactctctgtctcagacgga 2020
QY 1921 GGGCTCTGGGCTGGGAGTGCACCAATTTCTAGATAAAGATAAGAAAGGGAAGAAAGC 1980
Db 2021 ggtctctgggctgagggagtcgcgaatttcttagataaagaaaggggaaacacgc 2080
QY 1981 AAATACCGTCAATAATCTGTGATATGATATGATGAGAGTGCACGCAAACTTGTCTGAA 2040
Db 2081 aaataccgataaataatctgtgtgaltgctacgcgaagtgcaagcgcaaatctgtcgaa 2140
QY 2041 AACTTAATTAAGCTTTGGCTTTTGCAACTGTTGCTGTGCGATAAAGATTTCTTAGTGTCT 2100
Db 2141 aaacttaattagcttgctcttcttgcaactctctgtgtagcgataaagattctcttagtgc 2200
QY 2101 AAAATCATACTGATACCTATGACGAGAGCCCTTATATCCAAACATTTACAGATGTAGT 2160
Db 2201 aaaaatcatatcgtataactatgcagagagccctctatatccaaacatataagaaatgtagt 2260
QY 2161 GGGTTCATAGGTTGTCCTTAGATAAACTTCTGCTCTTGGAGTCAATTAACCCCTGTT 2220
Db 2261 ggttctcagtgctctcttagaataaactccgtgctcttgagctcaataaccctgct 2320
QY 2221 TTGAAGGCGACGCTGTTTAPGACGATGATGATGATGATGATGATGATGATGATGATG 2280
Db 2321 ttgaagaaggagagctcgtctatagccagctcaagtaagatctgaaagaaatataactgcg 2380
QY 2281 TATCTGAGGTGAAGAGGTTCTTGGGGAATATGCTTTTAAACATGATGTTGGAGCTTCT 2340
Db 2381 tatctcagtgagaaaggtctcttggtgggaataatagctttaacatgatgttggtgagctctc 2440
QY 2341 TCTCATTTCTTATCTGTAATACCTGATTTGTTTGTATACCTATGCTCCATACATCAACTG 2400
Db 2441 tctcatcttatactcgtaatcctgcgtatctgttgaataccctgctccatcatcaaaactg 2500
QY 2401 AATCTGACATATATAGCTGAGACAGCTTCCGAGAGAAAGTACAGAAAGATCTTTT 2460
Db 2501 aatctgacctataaagtcgagaaagctctcgcgaagaaagtaagaaagaaagctcttc 2560
QY 2461 GATGACAGCAACCTCTTCAATTTATCTTTGCTATAGGGGTGAAGTTTGAAGTTCTCT 2520
Db 2561 gatgacagcaacctctcattatctcttgccataaggggtgaaagtttggaagttctctc 2620
QY 2521 GATTGTAAATGACTTTTCTTATGATCTGACTTTATCTATGTCTCTGATCTTATCCGAAT 2580
Db 2621 gatgtgaatgactttcttatactgacttactatccatgttccgtctctatccgaaat 2680
QY 2581 GATCCCAATGACATACGACACTTTGTAATACGGGAGGCTCTTGAGAAATTTATGCCAAT 2640
Db 2681 gatcccaatgacctacagaaactgtatacagcgagagccctcttggtgaaactatgccaat 2740
|||||
```

Oy	2641	AACTTACGACAGACAGCCCTTGCAATGTCGTCAGAGCAGTCACTACGCTTCTCTATAG	2700
Db	2741	aacttaagacagacagcgaccttgcaatgctgcgcgagcagcactaagcctctctccatg	2800
Oy	2701	TTTAAAGGCGCCGGCCAGTTTGCTTTGAAGTTTCGGGATGCTCTCAGGATTTATATAGTA	2760
Db	2801	ttttaaagtgctgcgcgcagcttcgtcttgaaagtcgtgcgaccccaagatataaagta	2860
Oy	2761	GATCTTGGGGGTAAATTCCAAATTCATGAGGAGCGCTCTCTCATGTCTCAGAAATTCGT	2815
Db	2861	gattcttg9gg9gtaaagtccaattcttagagcgctctctcatgtctcagaattctg	2915
RESULT	4		
ID	AAx06828	standard; DNA; 3000 bp.	
XX	AAx06828;		
AC	AAx06828;		
DT	26-APR-1999	(first entry)	
DE	Chlamydia pneumoniae surface exposed protein Omp5 DNA.		
XX			
KW	Omp5; outer membrane protein 5; surface exposed protein; antigen;		
KW	Infection; diagnosis; vaccine; atherosclerosis; asthma; ss.		
XX	Chlamydia pneumoniae.		
OS			
XX			
FH	Key	Location/Qualifiers	
FT	CDS	259..3000	
FT		/*tag - a	
XX	MO9858953-A2.		
PN			
XX	30-DEC-1998.		
PD			
PF	19-JUN-1998;	98MO-DK00266.	
XX			
PR	23-JUN-1997;	97DK-0000744.	
XX			
XX	(BIRK/) BIRKELUND S.		
PA	(CHR1/) CHRISTIANSEN G.		
PI	Birkelund S, Christiansen G, Knudsen K, Madsen A;		
PI	Mygind P;		
XX	WPI: 1999-105610/09.		
DR	P-PSDB; AAW88429.		
XX			
PT	Species-specific test for identifying mammals infected with		
PT	Chlamydia pneumoniae - comprises detecting antibodies specific for		
PT	outer membrane proteins of C. pneumoniae or nucleic acids encoding		
PT	these proteins		
XX			
PS	Disclosure; Page 73-77; 115pp; English.		
XX			
CC	This DNA sequence encodes the novel surface exposed protein Omp5		
CC	(see AAW88429) of Chlamydia pneumoniae, a human respiratory pathogen.		
CC	It is described as a subsequence of a claimed nucleic acid fragment		
CC	(see AAX06817) encoding Omp5 (see AAW88418). The invention provides a		
CC	new species specific test for identifying mammals (including		
CC	humans) infected with C. pneumoniae. The test comprises detecting		
CC	antibodies specific for surface exposed proteins Omp4-Omp15 (see		
CC	AAW88417-28) or detecting nucleic acid fragments encoding them (see		
CC	AAW06816-27), especially by PCR. The proteins are also used in the		
CC	diagnosis of C. pneumoniae infection in mammals. The nucleic acids		
CC	and proteins can also be used in the immunization of mammals, the		
CC	nucleic acids being particularly useful as DNA vaccines for		
CC	effecting in vivo expression of antigens. The vaccines may also		
CC	prevent atherosclerosis and bronchial asthma, which are possibly		
CC	associated with C. pneumoniae.		
XX			

Seq	Sequence	3000 BP	867 A	597 C	658 G	878 T	0 other:
Query	Match	97.4%	Score 2742	DB 20	Length 3000		
Best Local Similarity	100.0%	Pred. No. 0					
Matches 2742	Conservative	0	Mismatches	0	Indels	0	Gaps
0							
QY	1	ATGAAATCGCAATTTTCCNCGTGTAGTCTCTTCGACATTTGGCATGTTTACTAGTGT	60				
DB	259	atgaaatcgcaatttcttcgtttagtgcctcttcgaaatgacatgcatgcttttactagt	318				
QY	61	TCCACTGTTTTGCTGCAACCTGCTGAAAATATAGGCCCTCTGATAGCTTTGACGGAAT	120				
DB	319	tccactgcttttctgacgaactctgcaaaaataagccctctgacatgcttgcaggaagt	378				
QY	121	ACTTAACAGCGAGACCATTCCTCTTAAATATAGACACTAGCGAATGACATATCTCGACA	180				
DB	379	acttaacacagcgacacctatactccctaaataagaaactcggaaatgaaactactcggaca	438				
QY	181	GGAGATATATACCTCTGCAAAACCTTGGGGATTCGGGACCTTTTAAAGAAAGGTTGTTTCT	240				
DB	439	ggagatatatactctgcaaaaccttggggttcggcagctttaaaggaaggttcttct	498				
QY	241	GACACTACGGAATCTTTAAGCTTTGCCGGTAAAGGGGTACTCATCTTTTAAATAT	300				
DB	499	gacactacggaatctttaagcttgcgcggttaagggttactcaacttctttaaatt	558				
QY	301	AGCTATAGTCGAAAGGCCGACACTTCTGTATCAACTGATAAAAATGTGCGTAAACA	360				
DB	559	aagctatagtcgaaaggcgcaagcaactctctgttaacaactgataaaatctgtcgttaaca	618				
QY	361	GGAATTTTCAGTCTTACTCTTTAGCGGCCCATCATCGGTATGCATCAACACCCCTCAGGA	420				
DB	619	ggatttctcgagcttactcttcttagcgcccatcatcgtatatacaaccccccttagga	678				
QY	421	AAAGTGCAGTTAAATGTGAGGGGATCTTACATTTGATTAACATGGAACATAATTTATTT	480				
DB	679	aaagtgcaagttaaatgtgagggaggtcttacaattgtatacaaatgaaactatattatt	738				
QY	481	AAACAAGATTAATGAGGAAATGGCGGAGCATTTCTACCAAGAATCTTCTTTGAAA	540				
DB	739	aaacaagattacttctgtaggaaaaatgvcggagccattctccaaagaattcttcttgaaa	798				
QY	541	AACACACGCGGATCGATCTTTCTTTGAAAGGAAATTAATCGACGCAACAGGAAAAAGGT	600				
DB	799	aacacgacgvggttcgaattcttcttgaaaggaataaactcgaagcgaacgvggaaaaaggt	858				
QY	601	GGGGCTATTTGCTACTGTACTGTACTGTATATTAACAATTAATAGCGCTCTACCCCTTTC	660				
DB	859	ggggctatttgcgtactgctactgtagtatatacaataataagcgtcctcctacccttc	918				
QY	661	TGGAACATATTTGCTGAGAGCTGACAGTGGAGCTATTAATAGCACAGGAACGTGTACATT	720				
DB	919	tggaaacaatattgctggaagctgcgaagtggaagctataaataagacaggaacgttacaatt	978				
QY	721	ACAGGGAATACGCTCTTATTTCTGAAAATAGGTACACGCGCACCCAGAGAAATGGA	780				
DB	979	acagggaatacagctcctgtatattcttgcgaataagtgtagcagcgaacggaggaatgga	1038				
QY	781	GGAGCTTTTTCGAGATGCCGATGTTTACCAATCTGGGAATCGAGAGTAACTTTTCTCA	840				
DB	1039	ggagctcttctcggagatgacgaagtgttaccatatctgggaatcagaagttaacttctca	1098				
QY	841	GGAAACCAAGCTGTAGCTATATGGCGGAGCCATTTATGCTAAGACGTTTACACTGCGTTC	900				
DB	1099	ggaacaaccaagcgttagctaagtcgagacattatgcttlaaagaacttacaactgcttcc	1158				
QY	901	GGGGGGGGGGGGGATGCTCTTTCTTAACATATAGTCCAGGATACCACTCAGGTAAAT	960				
DB	1159	ggggggggggggggtatccctcttcttaacaataagtcacaagtagtcaacgcgtcaggttaac	1218				
QY	961	GGTGAAGCATTTCTATACTGGACACTGGAGAGTATAGCTTTTCAGCAAGAACAGGGGAC	1020				



Db 1219 gctgagacatcttcatactgacgctgagagtgtagtcttccagcaagaagcgggac 1278  
 QY 1021 ATTACCTTCATGGGATGGCATGTTGTGCAACTACACCAAACTACAAAAGAAATTC 1080  
 Db 1279 attacctcaatgggaatgcgaatgtgtgcaactacaccacaatacaaaaagaattct 1338  
 QY 1081 ATTGACATAGGATCTACTGCAAAAGATCACGAATTTACGTGCATATATCTGGGCATATG 1140  
 Db 1339 attgacatagatctactgcaagaatcacgaatttaagctgcaatctcgtgcatagatc 1398  
 QY 1141 TTTTTCACGATCCGATCTACTGCTAATACGGCTGGGATTCCTACAGTACTTTAAATCT 1200  
 Db 1399 ttcttcacgataccgatactgatacagctgctgagattctacaga tacttcaatctc 1458  
 QY 1201 AATAGGCTGATGAGGTAAATAGTATAGTATAGTGGGCGATGTTTCTTTCGGTGA 1260  
 Db 1459 aataagagctgatacgaatagatacagatacagatgagctgcttcttcgagaa 1518  
 QY 1261 AAGCTCTGTAAGATGAAGCAAAAGTTGACAGAACCTCACTTCTACGCTGAAGCAAGCT 1320  
 Db 1519 aagctctgtaagatgagcaaaagtgcagacaacctcttaagctgtaagcagcct 1578  
 QY 1321 GTACTCTACTGACAGAAATTTAGTACTTAAAGCTGTCTACTCTCGATACGAAAGCC 1380  
 Db 1579 gtaacttaactgcaagaaattagtaacttaaacgtgtgtcaactctcga tactgaaagc 1638  
 QY 1381 TTTACTACGACCGGGGTTCTCTGTTATTATGGATCGGGACACAGCTTAAAGCAAGT 1440  
 Db 1639 tttaactcaagccgggttctctctgtatataatgagtcgggcaaaagcttaaaagcaagt 1698  
 QY 1441 ACAAGAGAGGTCACTTTAACAGGTCTTTCATTCCTGTAGACTCTTTTAAAGCGAGGTAA 1500  
 Db 1699 acagagagagctcaacttaaacaggtcttccatctctgtagactctttaggcgaggtgaa 1758  
 QY 1501 AAGTTTATATGTTGGTTCTGTCGACGAAGTAAATATAGCCCTTACTGTCGATCTT 1560  
 Db 1759 aaggtgtcaatgtgcgtcttcgacgaagtaaaaagtcaccccttaagtcgtgcgattctc 1818  
 QY 1561 CTTTGGATACCAAGGAATGCTTATGAATAATCACGACTTGAAGAAAACATCAAGACTT 1620  
 Db 1819 ctttggataccaagaagtaagtctatgaaatacgaacttagaaaaactcaagactct 1878  
 QY 1621 TCATTGTGACGCTCTGTCTGTGCTGCTGACTGCAACAATACAGATGTTCCAGCGGTCT 1680  
 Db 1879 tcaatttgtagcgtctctgtctcgtgactgcaacaactcaagtgctccagcgtctcct 1938  
 QY 1681 ACAGTAGCACTCTAGCACTATGAGTATCAAGGTACTGGGAATGACTTGGGTGAT 1740  
 Db 1939 acagtagcaactctcactacgactatggtaacaggtactctgggaaatgacttggtgat 1998  
 QY 1741 GATACCGACGACATCCCAAGACTAAGACAGCGCATTTAGCTTGACCAATATACAGCTTAC 1800  
 Db 1999 gataccgcaagcactcccaagaactcaagaacgacatlaagcttgagaccataacaggtac 2058  
 QY 1801 CTTCCGAATCCTGAGCGTCAAGGACCTTTAGTTCCTAATATAGCCTTGGGATCTTTTCA 1860  
 Db 2059 ctccgatactcgtgctgaagaccttagttcctaataagccttgggatacttttca 2118  
 QY 1861 GACATCCAAACGATTCAAGGTGATAGAGAGAGTCTTTGACTCTTTGTCAGATCGA 1920  
 Db 2119 gacatccaaagcattcaaggtgtaataagagaaggtgcttgactcttgltcaagatcga 2178  
 QY 1921 GGCTTCGGGCTCGGAGTGCACCAATTTCTAGATTAAGATAAGAAAGGAAAGCAAGC 1980  
 Db 2179 ggtcttcggctcgaggagtcgcgaattctttagataaagaaggggaaaaaacgc 2238  
 QY 1981 AAATACCGTCATTAATCTGTGTGATATGCTATGAGAGTGCAGCGCAACTGTGTTGAA 2040  
 Db 2239 aaataccgatacaatctgtgtgataatcgtgaggtgcagcgcaactgtgtctgaa 2298  
 QY 2041 AACTTATATGCTTGGCTTTTGCAACTTGTGTGATTAAGATTAAGATTTCTTAACTGCT 2100  
 Db 2299 aacttaattagcttgccttgccttgcgaactctctgtgtagcataaagaatttcttagtcgt 2358

QY 2101 AAAATCATCTGATACCTATGACGAGAGCCCTTCTATATCCACACATTTACAGATAGT 2160  
 Db 2359 aaaatcatatcigtactactatgcaagagcctctatataccacaacatatacagaatgagc 2418  
 QY 2161 GGGTTCTAGGTTGCTCTTTAGATAACTTCCTGCTCTTTGGAGTCAATAACCCCTGTT 2220  
 Db 2419 gggttcataagttgtctcttagataaactctcctgctgcttgagatcaaaacctgct 2478  
 QY 2221 TTGAAGGAGCGCTTCTTATGACGACGTAGTAAATGATGTAAGACAAATATACGCG 2280  
 Db 2479 ttagaaggagagctcgcttaagccacgtcagtaagatccgaaagaaatatacgcg 2538  
 QY 2281 TATCCTAGGTTGAAGGTTTGGGGGATTAATGCTTTTAACTGATGTTGGAGCTTCT 2340  
 Db 2539 tatccctgaggtgaaggtctcttggtgggataaagctttaacatgattggtgagctct 2598  
 QY 2341 TCTCATTTCTATCTGTAATACCTGACATGTTGTTGATACCTATGCTCATACATCAACTG 2400  
 Db 2599 tctcatcttatacctgataaccgcatgtgttgatccctatgtctccatacatcaaaactg 2658  
 QY 2401 AATCTGACCTATATAGTCTGAGACAGCTTCTCGGAGAAAGGTACAGAAAGATCTTTT 2460  
 Db 2659 aatctgacccatacagctcaggaagctcttcggaagaaagtaagaagaatcctct 2718  
 QY 2461 GATGACAGCAACCTCTTCAATTTATCTTTGCTATAGGGGTGAAGTTTGAAGTTCTCT 2520  
 Db 2719 gatgacagcaacccttcaattatcttgcctataaggtgaggttgaggttgcgtctc 2778  
 QY 2521 GATTGTAATGACTTTTCTTATGATCTGACTTTTATCCATGTTCTGATCTTATCCGCAAT 2580  
 Db 2779 gatgtgaatgactttcttataatgatacttaacatccatgcttccgtatccatccgaat 2638  
 QY 2581 GATCCCAATGACATACAGACACTTGTATATCAGCGACCTCTTTGGAAACTTATGCAAT 2640  
 Db 2839 gatcccaatgacataagcactatgataatcgaagcctcttggtggaactatgcgaat 2698  
 QY 2641 AACTTAGCAGACAGCGCTTGCAGATGCTGTCGACGAGCATACAGCTTATGCTCTATG 2700  
 Db 2899 aacttagcagacagccttgcaagtgctgcaagcagtcagtaacatgactctctccatg 2958  
 QY 2701 TTGGAATGCTCGGCCAGTTGTCTTTGAGTTCGTGAGTCCGTCGATCC 2742  
 Db 2959 ttggaatgctcgccagcttgcttcttgaaagtcgtgagatcc 3000

RESULT 5  
 AAA30850  
 ID AAA30850 standard; DNA: 2784 BP.  
 XX  
 AC AAA30850:  
 XX  
 DT 29-AUG-2000 (first entry)  
 XX  
 DE Chlamydia antigen CPN100635 gene open reading frame.  
 XX  
 KM Chlamydia antigen; diagnosis; infection; community acquired pneumonia;  
 KM therapy; upper respiratory tract disease; bronchitis; sinusitis;  
 KM asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;  
 KM ds.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..2784  
 FT /tag= a  
 FT /transl\_except- (pos:1078..1079; aa:Ser)  
 FT /transl\_except- (pos:1110..1111; aa:Glu)  
 FT /transl\_except- (pos:1790..1791; aa:Phe)  
 FT /transl\_except- (pos:2437..2438; aa:Gly)  
 FT /transl\_except- (pos:2613..2619; aa:Ala)  
 FT /partial  
 FT /product= Chlamydia antigen CPN100635



FT sig\_peptide /note= "no stop codon given"  
 FT 1..129 /\*tag= b  
 FT mat\_peptide 130..2784 /\*tag= c  
 FT /note= "mature CPN100635"  
 XX NO200032794-A2.  
 XX 08-JUN-2000.  
 PD 01-DEC-1999; 99MO-CA01147.  
 PF 01-DEC-1998; 98US-0110339.  
 PR 01-DEC-1998; 98US-0110340.  
 PR 01-DEC-1998; 98US-0110427.  
 PR 01-DEC-1998; 98US-0110428.  
 PR 01-DEC-1998; 98US-0110438.  
 XX (CONN-) CONNAUGHT LAB LTD.  
 PA Murdin AD, Oomen RP, Wang J;  
 XX MPI; 2000-412339/35.  
 DR P-PSDB; AAY90237, AAY90238.  
 XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for  
 PT preventing, diagnosing and treating diseases such as community acquired  
 PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset  
 PT asthma -  
 PT Claim 2; Page 146-147; 174pp: English.  
 PS This sequence encodes a Chlamydia antigen of the invention, designated  
 XX CPN100635. The nucleic acids (and their complementary sequences) may be  
 CC used as diagnostic agents for detecting the presence of nucleic acids  
 CC encoding Chlamydia antigens in samples according to standard methods,  
 CC and therefore, for diagnosing Chlamydia infections. For example, they may  
 CC be used as primers and probes for diagnostic polymerase chain reaction  
 CC (PCR) assays. Antisense sequences may be used to down regulate  
 CC expression of the proteins and may be used to treat infections. The  
 CC nucleic acids may also be used to produce the protein antigens they  
 CC encode according to standard recombinant DNA methodologies. The  
 CC proteins may then be used as antigens for the production of antibodies  
 CC (i.e. as vaccines) for preventing infection by Chlamydia. The  
 CC antibodies may also be used as diagnostic reagents for detecting  
 CC infections. Chlamydia is a pathogen implicated in the development of  
 CC (for example) community acquired pneumonia, upper respiratory tract  
 CC disease (especially bronchitis and sinusitis), asthmatic bronchitis,  
 CC adult-onset asthma and acute exacerbations of asthma in adults.  
 XX Sequence 2784 BP; 781 A; 571 C; 620 G; 812 T; 0 other;  
 SQ

Query Match 95.3%; Score 2683.4; DB 21; Length 2784;  
 Best Local Similarity 99.1%; Pred. No. 0;  
 Matches 2763; Conservative 0; Mismatches 16; Indels 10; Gaps 6;

QY 1 ATGAATCGCAATTTTCCTGGTAAAGTCTCTTCGACATTGGCATGTTTACTAGTTGT 60  
 Db 1 atgaatcgcaatttctcgttagtctctcttcgcattgagctgtttactagtctg 60  
 QY 61 TCCACGTGTTTGGTGAATCGTGAATAATATAGACACTGGAATAGACTTACTGTACA 120  
 Db 61 tccacgtgttttgcgtgaactcgtgaataatagagccctcgttagcttgaaggaagt 120  
 QY 121 ACTAACACAGGACCTTACTCTCTAAATAATAGACTGGAATAGACTTACTGTACA 180  
 Db 121 actaacacaggacactactactcctaaataatagactactggaataagactactactcgtaca 180  
 QY 181 GGAGATATTAATCTGTGAAAACCTTGGGANTGGGACGCTTAAAGAAAGGTTGTTTCT 240  
 Db 181 ggaatataactctcgaanaaccttgggattcggcagcgttaacgaaggttgttttct 240

QY 241 GACACTAGGAATCTTTAGCTTCCCGTAAGGGGTACTACTTTCTTTTAAATAT 300  
 Db 241 gacactaggaatcttttagcttcccgtaaggggtactacttttctttttaaataat 300  
 QY 301 AAGTCTAGTGTGAAGGGCGAGCACTTTCGTACAACTGATTAATAAATCTGCTTAACA 360  
 Db 301 aagtctagtggtgaagggcgagcactttcgtacaaactgattataaataatctgcttaaca 360  
 QY 361 GGATTTTCGAGTCTTACTTCTTTCAGGCGCCCATCATCATGATATACAAACCCCTCAGA 420  
 Db 361 ggattttcgagtcttacttcttcttcagcgcccatcatcatgattatacaaacccctcaga 420  
 QY 421 AAAGTGACGTTAAATGTGAGGGGATCTTACATTTGATTAACAATGAACTATTTATT 480  
 Db 421 aaagtgacgtttaaattgtgaggggatcttcatcttataacaatgaaactatttatt 480  
 QY 481 AAACAGATTTACTGTGAGAAATGGCGAGCCATTCTTACCAAGAACTCTTCTTGA 540  
 Db 481 aaacagatttactgtgagaaatggcgagccatttcttaccagaactcttcttga 540  
 QY 541 AACAGACGGGATCGATTTCTTTGAAGGAATAATCAGCGCAGCAAGAAAAAGT 600  
 Db 541 aacagacgggatcgatttctttgaaggataatcagcgccagcaagaaaaagt 600  
 QY 601 GGGGCTATTTGCTGCTAGCTGCTAGCTAGATATTACAAATATAGGCTCCTACCTCTTC 660  
 Db 601 ggggctatttgcctgactgctgactgactagatatttacaaatattaggctcctaccccttc 660  
 QY 661 TCGAACATATTTGCTGAAGCGCAGGTCGACTTAAATATACAGAACTGTACATTT 720  
 Db 661 tcgaacatatttgctgaagcgccaggtcgactttaaattacagaaactgtacaatt 720  
 QY 721 ACAGGAAATACGTCCTGTTATTTCTGAAAATAGTGTACAGCGACCGAGAAATGA 780  
 Db 721 acaggaaatacgctcctgttatcttctgaaaatagtgtacagcgacggagaaatga 780  
 QY 781 GGACTCTTTTGTGAGATGCCGATGTTACATATCTGGGAATCAGAGTGAATTTTCTCA 840  
 Db 781 ggaactcttttgtgagatgccgatggttacatattctgggaatcagagtgaaatttctca 840  
 QY 841 GGAACCAACGCTGATGTAATGGCGAGCCATTATGCTAAGAGCTTACACTGCTCTCC 900  
 Db 841 ggaaccaacgctgattgtaattggcgagccattatgcttaagagcttacctgctctcc 900  
 QY 901 GGGGGGGGGGGGATCTCTTTTCTAACAATATAGTCAAGGTACACTGACGATAT 960  
 Db 901 ggggggggggggattctcttttctaaacaatattagtcgaaggtacactgacgatatt 960  
 QY 961 GGTGAGCCATTTCTATACGACGCTGAGAGTGTAGTCTTTCAGCAGAGCAGGGGAC 1020  
 Db 961 ggtgagccatttctatacgacgctgagagtgtagtcttctcagagagcgaggac 1020  
 QY 1021 ATTACCTTCAATGGGAATGCCATTGTTGCAACTACACACAAACTACAAAAAATTC 1080  
 Db 1021 attaccttcaatgggaatgccattgtttgcaactacacacaaactacaaaaaatttc 1080  
 QY 1081 ATTGACATAGGATCTACT-GCAAAGATCAGAAATTTACGTGCAATATCTGGCAT 1139  
 Db 1081 attgacataggatctact-gcaaaagatcagaaattttacgtgcaaatattctggcat 1139  
 QY 1140 CTTTTCAGATCGAATTAATGCTAATAGCGCTGGGATTTCTACAGTATTTAAATCT 1199  
 Db 1140 cttttcagatcgaaatgaatgcttaattagcgctgggatttctacagatattttaaact 1199  
 QY 1200 CAATTAAGGCTGATCAGATATATAGATATATAGTGGGTGATTTCTTCTGTGA 1259  
 Db 1200 caatttaaggctgatcagatattatagatattatagtggttgatttcttctgtga 1259  
 QY 1260 AAAGCTCTGGAAGATGAAGCAAAAGTTGACAGACACTCTCACTTACGCTGAAGACCC 1319  
 Db 1260 aaagctcttgaagatgaagcaaaagttgacagacctctcaacttaccgctgaagacc 1319  
 QY 1258 aaagctcttgaagatgaagcaaaagttgacagacctcttaccgctgaagacc 1317





```

QY 1837 AATAGCCTTGGGATCTTTTTCAGACATCCAGGATTCAGAGTGTATGAGAGAAGT 1896
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1922 aatagccgtggtgtcttcttgatgtccatccatcagagctcatgacggagc 1981
QY 1897 GCTTTGACTCTTTTTCAGATCGAGGCTTGGGCTGGGAGTCCCAATTTCTTAAT 1956
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1982 acaagtcgtatctctcgtcaacaattgtggtatcagaatgcgagacttctgcat 2041
QY 1957 AAAGTAAAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2016
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2042 gaagatcagaagaagaacacagatgatatcgtatcagcgggtatgacatcaga 2101
QY 2017 GGTGAGGAGCAACTTTTCTGAAATTAATTAATTAATTAATTAATTAATTAAT 2076
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2102 ggaagattcttcacggtctcgaaattcttaatttcttcttcttcttcttcttctt 2161
QY 2077 AGCGATAAAGATTTCTTATGCTGCTTAAATTAATTAATTAATTAATTAATTA 2136
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2162 tacgacaaagacacatctgtgtcctaagaacacacacacacacacacacacac 2221
QY 2137 ATCCAAACACATTAAGAGATAGTGGGTTCAAGGTTGCTCTTAAGATTAATTA 2196
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2222 tacgacacccctcgagagctca-----agacctcgttaagatttctcagga 2269
QY 2197 TCTTGAAGTATAAACCCCTCTTTTACAGAGGAGCTGCTTATAGCCAGTCAAT 2256
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2270 aattcgatccctacatcttcttctcattcgtctgttcttcttcttcttcttct 2329
QY 2257 GATCGAAGACAAAGATTAACGCTGATCCGATCCGAGAGAGAGAGAGAGAGAG 2316
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2330 aacatgacacaaagatacagcgtatctcctcgtcctaagagagcgtggaatgagcc 2389
QY 2317 TTTTACATGATGTTGGAGCTTCTTCTCATTTCTTATCTCTCTCTCTCTCTCT 2373
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2390 ttctgtatagaatgtgagggagcctatcccgtagtctgttctgaagacgctgtcttg 2449
QY 2374 GATACCTATGCTCCATATACATCAAACTGATGACCTATATACGTCAGGACGCT 2433
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2450 gataccacacacacacacacacacacacacacacacacacacacacacacacac 2509
QY 2434 GAGAAAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2493
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2510 gaaacagcagacagagagcgtctctccaaagtgaagacccctcctcagcgtctct 2569
QY 2494 ATAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2553
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2570 gttaggataaaattgtgaattctcgcgata-----agctcagtagatcctccata 2623
QY 2554 TCTTATGCTCTGATCTTATCCGCAATGATCCCAAAAGCTACAGACCTGTATTCAG 2613
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2624 gcttaagtcctcgtatgtatcgtatcagatcagcgtcagacacacacacacac 2683
QY 2614 GGAGCCTCTTGGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2673
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2684 gggagatctctgtcagatgtgatacagcttctcagacacacacacacacacac 2743
QY 2674 GGCAGCTACTAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2733
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2744 ggaatcatcatgtccttctcctcctcctcctcctcctcctcctcctcctcct 2803
QY 2734 CGTGAGCTCTACAGGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2786
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2804 cgaagctctctcgttagctatgctatcagatcgtgaggaagattcgcgatttta 2856

```

```

RESULT 7
ID AA06821 standard; DNA: 2757 BP.
XX AA06821:
XX AA06821:
XX 26-APR-1999 (first entry).
XX

```

```

DE Chlamydia pneumoniae surface exposed protein Omp9 DNA.
XX
XX Omp9: outer membrane protein 9; surface exposed protein; antigen;
KW infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX Chlamydia pneumoniae.
XX
XX W0958953-A2.
XX
XX 30-DEC-1998.
XX
XX 19-JUN-1998: 98W0-DK00266.
XX
XX 23-JUN-1997: 97DK-0000744.
XX
XX (BIRK/) BIRKELUND S.
XX (CHR/) CHRISTIANSEN G.
XX
XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX Mygind P;
XX
XX WPI: 1999-105610/09.
XX P-PSDB: AAM06822.
XX
XX
XX Species-specific test for identifying mammals infected with
XX Chlamydia pneumoniae - comprises detecting antibodies specific for
XX outer membrane proteins of C. pneumoniae or nucleic acids encoding
XX these proteins
XX
XX Claim 6; Page 55-56; 115pp; English.
XX
XX
XX This DNA sequence codes for the novel 96.7 kDa surface exposed
XX protein Omp9 (see AAM88422) of the human respiratory pathogen
XX Chlamydia pneumoniae. By generating antibodies against C.
XX pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
XX was obtained which reacted with outer membrane proteins. The
XX antibody was used to identify the genes (see AAM88417-28) encoding
XX Omp4-Omp15 proteins (see AAM88417-28) in an expression library of
XX C. pneumoniae DNA. The genes are situated in 2 gene clusters:
XX Omp12, 11, 10, 5, 4, 13 and 14 in one cluster and Omp6, 7, 8, 9 and 15 in
XX the other, and encode polypeptides of about 89.6-100.3 kDa and
XX about 56.1 kDa. The invention provides a new species specific test
XX for identifying mammals (including humans) infected with Chlamydia
XX pneumoniae. The test comprises detecting antibodies specific for
XX Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
XX membrane proteins, especially by PCR. The proteins are also used
XX in the diagnosis of C. pneumoniae infection in mammals. The
XX nucleic acids and proteins can also be used in the immunization of
XX mammals, the nucleic acids being particularly useful as DNA
XX vaccines for effecting in vivo expression of antigens. The
XX vaccines may also prevent atherosclerosis and bronchial asthma,
XX which are possibly associated with C. pneumoniae.
XX
XX Sequence 2757 BP; 770 A; 564 C; 642 G; 781 T; 0 other;
XX
XX
XX Query Match 23.8%; Score 670.6; DB 20; Length 2757;
XX Best Local Similarity 55.8%; Pred. No. 9.2e-181;
XX Matches 1570; Conservative 0; Mismatches 1159; Indels 84; Gaps 12;
XX
XX
XX 1 ATGAAATCGCAATTTTCCGTGGTATGCTCTCTTCGACATTCGATGTTTACTAGTTGT 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 atgagatcgtcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 60
QY 61 TTCACGTGTTTGTCTGCAACGCTGGAATATATAGCCCTCTGATAGTTTGAGGAGAT 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 agtcttctcagatgctgctcagatccacatcagtaggagctgtagcagttataatggtat 120
QY 121 ACTAAGACAGGACCTATATCTCTTAAATTAATTAATTAATTAATTAATTAATTAAT 171
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 acaagcaccacagaaatttacccttaagcggcaactctcgtatgtagcagacactat 180
QY 172 ACTCTGACAGAGATATATATCTCTCAAAACCTTGGGATTCGAGCAGCTTAAACGAGGT 231

```

Db 181 attccgattggygatctcgataagccaagcagggaaacaaagagcttaacacaagt 240  
QY 232 TGTCTTTCTGACACTACGGAATCTTTAAGCTTTGCCGTAAGGGGTACTCACTTTCTTT 291  
Db 241 tgttttctaacacgcaggaatacttactcttaaggaaacgagattctcttccatctt 300  
QY 292 TTTAATATTAGTCTAGTGTGAAGGGCGACACTCTTCTGTACAACTGATAAAAATGTG 351  
Db 301 gacaaataattatctgctact---gtgcagagtgltgttctatgcaatacagcagctct 357  
QY 352 TCGCTAACAGAGATTTTCGAGTCTTACTTTCTTAAGCGGCCCATCGGTAAATCAACAC 411  
Db 358 gggattacgaattctcagagatttcaactctcagatcgctgagcctcctcaga----- 412  
QY 412 CCCCTCAGAAAAGGTGCGATTAAATGTGAGGGATCTTACATTTGATTAACAAATGGAAGT 471  
Db 413 -cccaaggtlaaagagcattcaaaattaccgattgctggtgttggagatagggat 471  
QY 472 ATTTTATTTAACAAGATTACTGTGAGAAAATGCGGAGCATTTTACCAAGAAATCTT 531  
Db 472 cttaacaaaatgaaaatgctctcagtgaaaatgggagccatcaaalacgaagaatttg 531  
QY 532 TCTTTGAAAAACAGCAGCGGATCGATTCTTTTGAAGGAAATTAATCGAGCGCAACAGG 591  
Db 532 tctttgactggagtagcgggtttgtagcgtctccttgcaatagctgctc-----g 582  
QY 592 AAAAAAGTGGGCTATTGTGCTACTGCTGATCTGATATTACAAATTAATAGGCTCTCT 651  
Db 583 caaaagggagggagctactatgcttctgtgactctgactctgtatcttctgaaagcaggaatc 642  
QY 652 ACCCTCTTCTGACAAATATTGTGAAAGCTGACAGTGGAGTATTAATATGACAGGAAC 711  
Db 643 ttgagcttcggaaaacaacatgctcagcaacatcaggaagcgcagctcctcgtcgaaggac 702  
QY 712 TGTACAAATACAGGGAATACGCTCTTGTATTTTCTGAAATAGTGTGACAGCAGCCGA 771  
Db 703 ctgtgacttcacaataccaataatcttcttgatgctgcaagcaactacaatagc 762  
QY 772 GGA-----AATGAGAGAGCTTTTCTGAGATGCCGATGTTACCATTTCTGG 819  
Db 763 ggaactatgtatgtlaacaaagcagggcgaaacccactatcttgaactcttcagga 822  
QY 820 AATGAGATGATTAATCTTCTGAGAAACCAAGCTGTAGCTAATGCGGAGCCATTATGCT 879  
Db 823 aatgagagcctgcatcttctgaaatacagcaggaatacagtgagtgagcttatacc 882  
QY 880 AAGAAGCTTACACTGCGTTCGCGGGGGGGGGGGGTATCTCTTTGTAACAAATATATC 939  
Db 883 aaaaattgtgtatctctcagagcagggagtgatatttcttaacaaagcctgag 942  
QY 940 CAAAGTACACGTGAGTAATGTGAGCCATTTCTAATCTGAGCAGTGTGAGTGTAGT 999  
Db 943 aatgcatctc-----taagggagggcaattgcgattctgaattctgagagattagc 996  
QY 1000 CTTTCAGAGAGAGGAGGAGATTAATCTTCATGGAATGCCATTTGTGCAACTACA--- 1056  
Db 997 attctcgaagatctcggcaatatacttctgagggcaactcagagcactcagcaggaat 1056  
QY 1057 CCACAAACTACAAAAAATTTCTATTGACATAGATCTACTGCAAAAGATCACAAATTTA 1116  
Db 1057 ccttgaggtgtgacagaaatgctatagatcttgatcagaaatgaaatlttataactc 1116  
QY 1117 CGTGCAATATCTGGGCTATGATCTTTTCTGACATCCGATTACTGCTAATAGCGCTGGC 1176  
Db 1117 cgaagcagctcgggaaataagttatttctatgctc-----ctatcagagctca 1167  
QY 1177 GATCTCAGATCTTTAATATCTCATTAAGGCTGATGAGGATATATGATATAGT 1236  
Db 1168 ggaagctactgtaagctctcttctgaaataagcgtgacgagagatctggaaatacctagaa 1227  
QY 1237 GGGTGAATGTTGTTTCTTGTGTAAGACCTCTGAAAGATGAAGCAAAAGTTGACAGAAC 1296  
Db 1237 GGGTGAATGTTGTTTCTTGTGTAAGACCTCTGAAAGATGAAGCAAAAGTTGACAGAAC 1296

Db 1228 ggtacatcgctttctctcgtgagagaactctcagaagaaggaacttaagaacctgacat 1287  
QY 1297 CTCATCTTACGCTGAGACACGCTGTAACTGTAACTGACAGAAATTTAGTAAACGT 1356  
Db 1288 ctgaagctacatttacaacagcgttgtagagcttctgcaagtgctgacttagtaatgaagat 1347  
QY 1357 GGTGCACTCTGATACGAAAGGCTTTACTGACACCGGGGGTCTCTGTTAATATGAT 1416  
Db 1348 ggaagtactgtagtgcacaataactataaagcagctcgaagtcgaagatcgatagat 1407  
QY 1417 GCGGCGACAAAGTTAAAAGCAAGTACAGAGAGGTCACTTTAAGAGTCTTTCAATTCCT 1476  
Db 1408 ggaaggaactacttttgagcgaagcgtcgaagggttactctcactcaatgagcctatcaat 1467  
QY 1477 GTAGACTTTTAGCGAGGCGGTAAAGAACTGTAAATGCTGCTTCTGACGCAAGTAAAT 1536  
Db 1468 atagattccttagatggtgacaataaagctcatcttaagcgaagcgaagaatagat 1527  
QY 1537 GTAGCCCTTAGTGGTCCGATCTTCTTTGATTAACCAAGGGAATGCTTATGAATAATCAC 1596  
Db 1528 gtgccttataagggcctatactgtagatgctcaggggaactatataagcatcat 1587  
QY 1597 GACTTAGAAAAACTCAAGACTTTTTCATTTGTGAGCTCTCTGCTGCTGATGCAACA 1656  
Db 1588 aatctcagtaacagcaggtcttctccttaataagacttctgcaagaagagatgact 1647  
QY 1657 ACTACAGATGCTCCACGCGTTCTACAGTACCAACTCTACGCACTATGAGTATCAAGT 1716  
Db 1648 actcaagatatacccgatacccaattctaaataactcgaatcactatggtatcaaga 1707  
QY 1717 ACTTGGGGAATGATGTGGGTGATGATACCGCAAGCACTCCAAAGACTACACAGCGCA 1776  
Db 1708 actggaataattgttctggtgcagagc-----aactgcgaataacaaaatgctacc 1761  
QY 1777 TTACCTTGACCAATACAGGCTTACCTTCGAATCCTGACGCTCAAGACCTTTAGTTCT 1836  
Db 1762 ttaacttgacttaaaacagagatataagcggaatccagaagctcagagacttctgctct 1821  
QY 1837 AATAGCCTTTGGGATCTTTTCAAGACATCCAGCGATTCAGAGTGTATGAGAGAACT 1896  
Db 1822 aatagcctggtgtctcttctgtagtctcgcctcaactcgaagcctatgagcagggc 1881  
QY 1897 GCTTGACTCTTTTGTGAGATCGAGGCTTCTGGGCTGGGAGTCCCAATTTCTTAGAT 1956  
Db 1882 acaagctcgtatctctgcatacaaatltggtgtatcaagaaatcgagacttctgact 1941  
QY 1957 AAGATTAAGAAAGGGGAAAAACGCAAAATACCGTATTAATGTGGTGAATATGCTATGGA 2016  
Db 1942 gaagatacgaagaagaaacaaacagtagtatactcatctcagcgggttatagcatagga 2001  
QY 2017 GGTGACAGCGCAACTGTGTTGAAAACTTAATTAATAGCTTTGCTTTGSCCACTCTTGGT 2076  
Db 2002 ggaagattcttcaacggtctcgaanaatttcttaatttctgttcttctgcaagcttctggc 2061  
QY 2077 AGCATTAAGATTTCTAGTCGCTAAATATCATATCTATACCTATGACAGAGCCTTCTAT 2136  
Db 2062 tagacaagaagcaactctgtgtgaaagacacataccatgtatataggaaggaatagat 2121  
QY 2137 ATCCAAACATATTAAGATGATGAGTGGGTTCATAGTTGTCTCTTAAGATTAACCTCTGGC 2196  
Db 2122 taacgaacactcggagagctca-----agacacctcgtaagatatttctgagga 2169  
QY 2197 TCTTGGAGTATTAACCCCTCGCTTTTGAAGGGCAGCTCGCTTATACCAAGTATAT 2256  
Db 2170 aatctcagctccctacacttctgtcttcaatgctcgttctttagcgcatacagcaaat 2229  
QY 2257 GATCTGAAGCAAAAGTATACGCTATCCCTGAGGTGAAGGTTCTTGGGGAATATAGCT 2316  
Db 2230 aacatgcacaagaatacaggtcattctcctgttaaggaagcctggggaataagatgccc 2289  
QY 2317 TTTAACAATGATGTTGGAGCTTCTTCTCATTTCTTAATCT---GAATCTCGCATGTTT 2373  
Db 2290 ttcgtagtagaagtgtgagagactaccggtagtgcttcaggaagctcgctctgggtg 2349





Oy	811	ATPCTGGGATCGAGGATGTAACCTTCTCGAAGAAACCAACAGCTAGTATGATGCGGAGGC	870
Db	107855	ATCTCTGGAAATATAAAGCTGACCTTGCCGAGAACTCTTCACTAATCTCAAGCGGAGCA	107796
Oy	871	ATTATGCTAAGAGCTTACACTGGCTCCGGGGGGGGGGGATATCTCTTTCTTAC	930
Db	107795	ATCTGTGCCATGTGTATAGATCTTTCCGGTGTGCGCC-----TACCCATTTTCAAA	107742
Oy	931	AATATATGCCAAGTACCACCTGCAAGTAAATGATGAGCCATTTCTATCTGCGAGCTGGA	990
Db	107741	AATGATATGCGGGAAACACAGCTGCGAGGCAAGGGGGCGCTATTCCATTTGCCGACTGTGGA	107682
Oy	991	GAGATGATCTTTACAGAGAAGCGGGAGCAATTACCTTCAATGGGAATGCAATGTGGA	1050
Db	107681	TCTTTAATCTCTCTGCAATTCMAAGGAGACATCACGTTCTTGGCAACACTTAACCTCA	107622
Oy	1051	ACTACACCAACAACTACAAAAAGAAATTTGATGACATAGATCTACTCGMAAGATCAG	1110
Db	107621	ACCTCCGGGCCAAMCATGACACGGAATGCTATCTACCTGGATGTCAGCAAAATTAAG	107562
Oy	1111	AATTAGGTGCAATATCGGGCATAGACATCTTTTCTAGCATCCGATCTCTGTATATAG	1170
Db	107561	AACTTAAGGCGACGCCCAAGCCATCTATATTTCTATCTATGATGATGATTAACACC	107502
Oy	1171	GCTCGGATTTCTACAGATTACTTTAAATCTCAATAAGGCTATGCAAGTATATGATACAGAT	1230
Db	107501	ACAGAG- - -CTTCAGACGTTTCTGACATCAACCAACCGATATGCAATCGCCTTAGAT	107445
Oy	1231	TATAGTGGGTGATGTTGTTTTTCTGGTGAAGAACTCTCTAAGATGACGAAAGTTGCA	1290
Db	107444	TATTCAGGAAGATGTAATTTTCTGGGGAAACCTCTGCAATAGAGGAAAGGTGCT	107385
Oy	1291	GACAACTCACTTTACGCTGAAGCAACCTGTAACTCTAATCTCAGAAATTTAGTACTT	1350
Db	107384	GATACCTTCACATCTATATTAAGCAACCATTTGGCTGTACCTCTGAACTTAAGCACTC	107325
Oy	1351	AAACGTGTGTCACTCTCGATACGAAAGCTTTACTCAGACCGGGGTTCTCTGTATTT	1410
Db	107324	AAAGGAATGTGCGATGATGTCMAAGGTTTTCACACAGCTGAAGGCTTCACTCCTC	107265
Oy	1411	ATGATGCGGGCAACAGTTAAAAAGCAAGTACAGAGAGTCACTTTAACGCTTTGCC	1470
Db	107264	ATGCACACGAGCAACAAAGCTCAAGGCAAGATCTGAACCTATCAAGCTTTACCAACTGTTC	107205
Oy	1471	ATTCCTGTAGACTTTTAGCGGAGGGGTAGAAAGTTGTATTTGCTCTCTTGCAGCAAGT	1530
Db	107204	GTTGATCTTTCTGCTCTTAGAGGGAAATATAGATGTGTCATTAACAGAGAGCAAC	107145
Oy	1531	AAAAATGTAGCCCTTATGGTCCGATTTCTTTTGATATACCAAGGAATGTTATGAA	1590
Db	107144	AAAACTATAACTCTAACCTCTCTCTTGTTTTCCAAATATGTAGCGGCAATTTTATGAA	107085
Oy	1591	AATCAGCACTTAGGAAAAACATCAAGATTTTCAATTTGTGAGCTCTCTCTGCGTACT	1650
Db	107084	AGGCATACGATTAAMCAAGGCTTCAGCAGCAACCTTTGGTGTATTACTGTGCTACTGCT	107025
Oy	1651	GCAACAACTACAGATGTTTCCAGGGGTTCTTACAGTACCACTCTTACGCACTATGSGTAT	1710
Db	107024	GCTTGCGCATTTTATATACGATGCGGCTTCTCACTTCTCCAGTACAAACTCCAGAACTCAT	106955
Oy	1711	CAAAGTACTTTGGGAATGACTTGGGTTGATGATACGCGCAAGCACTCCAAAGATATAGACA	1770
Db	106964	TACGGGATATCAGGA- - -CATTTGGGAAGCACTTGGGCGACACATCACTCAAAATTA	106908
Oy	1771	GCGACATTAAGCTTGACCAATACAGGCTACCTTCCGAATCTGAGGCTCAAGACCTTTA	1830
Db	106907	GGAACTATGACTTGGTAACTACGGGGCTCAACCCATATCTGAGGCTGAGGTTCCGTA	106848
Oy	1831	GTTCTCTAATAGCCTTTGGGATCTTTTTCAGACATCCAAAGCAATTAAGGTGATAGAG	1890
Db	106847	GTTCCCAATTCATATATGGGATATCTTATACGAAATTCGCACTCTTACAGAGATCAGACA	106788

QY	1891	AGAGTGGTTTGACCTTTGTCATGATCGAGAGCTTCGGGGGGGGGAGTGGCAATTTC	1950
Db	106787	TCCTCAGGGAATAGATATCTATACGACCAAGAGACTCTGGGATATGGAATTC	106728
QY	1951	TTATGATTAAGATTAAGAAAGGGGAAAAAGCAATACCGTCATTAATTCGGTGTATGCT	2010
Db	106727	TTCCATTAAGGATTAATATCAGAAACATACCAAGCATTCGACATTAATAACTTCGCGTATATT	106668
QY	2011	ATCGAGGTCGACGGCAAACTGTGTCGAAAACTTAATTAAGCTTTGCTTTGGCAAATC	2070
Db	106667	GTTGGAGGAAGTGGCAAGATTTTCTCGAAAAATCTCTAGGTAGCTTCTTCGCACTC	106608
QY	2071	TTTGGTACCGATAAAAGATTTCTTAGTGGCTAAATAATATACTGATACCTATCGACGAGCC	2130
Db	106607	TTTCGGTAAAGATTAAGAACCTTTTATAGTTGAAAAATCTCTCTCAATATTAAAGCTCG	106548
QY	2131	TTCTATTATTCACAAACATATACAGAAATGATAGTGGGTTCATAGTTGTCTCTAGATAAATT	2190
Db	106547	CTATTAACCTGCAACATCCAGACATTTCTCTAGAGAGCTTCCCATGCCCTATTTGGAAATATC	106488
QY	2191	CTGGCTCTTGGAGTCATAAACCCCTCGTTTTAGAAGGGACGCTCGCTTATAGCCACATC	2250
Db	106487	AACGCATGCTGAAAAAGATATTCCTCTCATTTTGAATGCCAGCTAAAGCTACAGCTACAT	106428
QY	2251	AGTATGATCTGAGACAAATATATACGCTCGTATACCTGAGTGAAGAAAGTTCTTGGGGAT	2310
Db	106427	AAAAATGATATGATGATACCTGCTATATCTTCATCTCTTAAGCTCAAGGCTCTTGGAACAT	106368
QY	2311	AATGCTTTTAAATCATGATTTGGAGAGCTCT--TCGTATTCTTATCTGTAATACCTGAT	2367
Db	106367	AACCTGTGGGCTCTAGACCTGGAGAGATCTGTGGCTATATATCTCCCTAAGAAGCACCG	106308
QY	2367	TGTTTTGATACCTATGCTCCATACATCAAACTGAATCTGACCTATATATAGCTCAGGACAGC	2427
Db	106307	TTCTTCCAGGAGATTTTCCCTCTTAAAGTTCGAGCAGCTCTACAGCCGCAACAAAC	106248
QY	2428	TTTCGAGGAAGATGACAGCAAGAAATCTTTATGATACAGCAACCTCTCAATTTATCT	2487
Db	106247	TTTAAAGAGATGGGGCTGAAAGCCGGCTGCTTTTATATATGAGACCTAGAGAACTGCTT	106188
QY	2488	TTGCCTATAGGGGTGAAGTTGAGAAAGTCTCTGATTTGATGACTTTCTTATGATCTG	2547
Db	106187	ATCCCTGTCCGCACTTCGGTTAGAAAAATCTCCGAAGTGAATAAATAATTTCGAGATT	106128
QY	2548	ACTTTATCTATGTTCTGATCTTATCCGCAATGACCAATGCAATGCAACGACATCTTGA	2607
Db	106127	TCTCTAGGCTCATTTGGATGATGTATCGTAAATAATCCGGTTCGCTACTCTCTAATG	106068
QY	2608	ATCAGCGGAGGCTCTGGGAACCTTATGCCAATATCTTAGCAGACAGCGCTTGCAAGTG	2667
Db	106067	GTCAGTGGAGGCTCTTGGAATCTTGCCTATGTAAAAACCTCGACAGCAAGGCTCTTATGCA	106008
QY	2668	CGTGCAGGCACTACATAGGCTTCTCTCTCTCTATGTTTGAAGGCTCGGCGCAGTTTGCTTT	2727
Db	106007	AGTGTGGAAACCACTGTACCTCTCTCCCTCATATGTAAACTCTCTGGGGAGAGCTGCTAT	105948
QY	2728	GAACTGTGTGATCTCAAGGATTTATATATGTATAGATCTTGGGGTAAAGTTCCAAATCTAG	2787
Db	105947	GAGCTTCGTGCTCAGCAACATCTACAAATGTAGATTGTGGGCTAAGATCACTATCTAG	105888
QY	2788	GAGGCTCTCTCATGTCTCA 2806	
Db	105887	TTTCTTACTTTCTCTCCCTAA 105866	

RESULT	9
AAx91990	
ID	AAx91990 standard; DNA; 1230025 BP
XX	
AC	AAx91990;
XX	
DT	13-SEP-1999 (first entry)
XX	



DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope; ss.

OS Chlamydia pneumoniae.

XX MO9927105-A2.

PD 03-JUN-1999.

XX 20-NOV-1998; 98WO-1B01890.

XX 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (GENST ) GENSET.

XX Griffiths R;

DR WPI: 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

PT Claim 1; Page 291-611; 1912pp; English.

CC The present sequence represents the complete genome of Chlamydia  
CC pneumoniae, and encodes proteins AAY34584-y35879. C. pneumoniae causes  
CC respiratory disease such as pneumonia and bronchitis and is thought  
CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,  
CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
CC encoded by the open reading frames of the C. pneumoniae genome (see  
CC AAY34584-y35879) can be used in immunogenic compositions as vaccines.  
CC Vectors containing C. pneumoniae nucleotide sequences can also be  
CC used as immunogenic compositions, especially where the vector directs  
CC the expression of a neutralising epitope of C. pneumoniae.

CC Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match 23.1%; Score 649.4; DB 20; Length 1230025;

Best Local Similarity 54.5%; Pred. No. 2.3e-173;

Matches 1548; Conservative 0; Mismatches 1231; Indels 60; Gaps 10;

QY 1 ATGAAATGCAATTTTCTGTTAGTCTCTCTGACATGTCATGTTTACTAGTGT 60

DB 507718 atgaaataacccttgcacaactcctgactctcttgcacctctgcaatccatctatg 507777

QY 61 TCCACTGTTTTGCTGCAACTGCTGAAAATATAGGCCCTCTGATAGCTTGAACGA 117

DB 507778 agcatcgcaactacgagcagatgctctcttaccctacagatgacttgatgagcg 507837

QY 118 ACTACTAACACGACCCCTACTCTTAATAATAGACACTACGATTAACACTACTG 177

DB 507838 ggcggcctcacttacttcaacaaatcacaagatgccaatgcaagaacatgctcta 507897

QY 178 ACAGAGATATATCTGCAAAAACCTTGGGATTCGCGACTTATGAGAAGGTTT 237

DB 507898 tcaggaatagctataataacagatgctggaagaaacacgacatlaacagcgctgctt 507957

QY 238 TCTGACTACGGAATCTTTAAAGCTTTGCCGGTAAAGGGTACTACTCTTTTAAAT 297

DB 507958 acagaactacagggcgatctgacttactggaagsgatcattcttcatacaacag 508017

QY 298 ATTAAAGTCTAGTGC---TGAAGCGCAGACATTTCTGTTACAACTGATTAATAATCTGTG 354

DB 508018 gtagatgctgggttcgaatgcaagatgctgcaacacacatgctgataagcccttaaca 508077

QY 355 CTAAACAGATTTTTCGAGTCTTCTTATAGCGGCCCTCATCGGTATTCACAAACCC 414

DB 508078 ttcaacagatttctcaaccttctcatctgacgctc-----ctgaaactaacagtgtct 508131

QY 415 TCAGAAAAGTGCAGTAAATGTCGAGGGAGCTTACATTTGATTAACATGAACTATT 474

DB 508132 tcaggaataaagtaacttaagctctgcaagccttaactacacgataatggaagct 508191

QY 475 TTATTTAAACAGATTACTGTGAGAA-----AATGGGGAGCCATTTCTTACCAAG 525

DB 508192 cctttagccaaaacgctcccaatgagtaataacaatgcygagcgatcaccgcaaaa 508251

QY 526 AATCTTTCTTTGAAAACGACGGGATTTCTTTTGAAGGGGATTAATCGACGCA 585

DB 508252 actcttccattcttgcggaatccctcttataaccttcactagtaataatgcgcaaaa 508308

QY 586 ACAGGAAAAAGTGGGGCTATTGCTACTGCTACTGATATTAATTAATTAACG 645

DB 508309 -----aaatggtgagcgatctctatagctctgcgcygcaagatctcaggaaccc 508362

QY 646 GCTCTTACCTCTTCTCGAACAATATTGCTGAAGCTGCGAGGTGAGACTTTAATAGACA 705

DB 508363 ggcagtagtcttcttataataaagaaactggyggtggtggtctggttgaa 508422

QY 706 GGAACGTATACATTTACAGGGAATACGCTCTGTTATTTCTGAATAATGACAGCG 765

DB 508423 gccagctcctcgattactcaataatagctcccttcttcttctgcgaaacatgcaagat 508482

QY 766 ACCGAGAAATGAGAGAGCTCTTCTGAGA-----TCCGATGTTTAC 810

DB 508483 gctgcagcaaggcggygcatttattgtgaaaaaacagagagagactctacttact 508542

QY 811 ATATCTGGGAATACAGGTACTTCTTCTGAGAAACCAAGCTGTAGCTATATGGCGAGCC 870

DB 508543 atctcgyaaataaagatcgtgacttccgcggaactcttcagtaactaaagcgagga 508602

QY 871 ATTATGCTAAGAGCTTACACTGCTCCGGGGGGGGGGGATTCCTCTTTCTAAC 930

DB 508603 atctggtcccaatggtctatgacttcttccgtctggtcc-----tacctatttcaat 508656

QY 931 AATATAGTCCAAAGTACCACTGCAAGTAATGTGAGCCATTTCTATAGTCGACGTGA 990

DB 508657 aatagatgcyggaacacagctcagcgaagcggygcatttgcaattgcgactctga 508716

QY 991 GAGTGTAGCTTTTCAAGCAACAGGAGCATTTACTCTTAATGGGAATGCCATTTGTGA 1050

DB 508717 tcttaagctctctcgaataacaaagagacaacagcttcttgcgaatacttaacctca 508776

QY 1051 ACTACACCAACATCTCAAAAAGAAATTTGATGATAGTACTACTCAAAAGATCAGC 1110

DB 508777 acctcgcgcaacatcgcaacgaaatgctatctactcgtggtctgcagcaaaaattag 508836

QY 1111 AATTTAGTGCATATCTGGCATAGACATCTTTTCTACGATCCGATTTACTGCTAATAC 1170

DB 508837 aacttaagggcgcccaagccaatctatctattctatgcagatgcgattcactaacacc 508896

QY 1171 GGTGGGATTTACAGATCTTAAATCTCAATTAAGCGTGAAGGATTAATAGTACAGAT 1230

DB 508897 acaagag---cttcaagcttctgcacatcaacacacggaataagcaactcgctttagt 508953

QY 1231 TATAGTGGTGCATGTTTCTTCTGTTGAAGCTCTCTGAGATGAACCAAAAGTTGCA 1290

DB 508954 tattcgaagaagatgtattcttctggygaagcctctctgcagatgaagcgaagcgtgt 509013

QY 1291 GACAACTACTTCTTACGCTGTAAGCAGCCTTATCTTACCTGACGAGAAATTAAGTACT 1350

DB 509014 gataacttcaactatataataaagcaacatggtctcagctctcgaaccttcaacacc 509073

QY 1351 AAACGTGTGTACACTCTGATACGAAGGCTTTTACACAGCGGGGTCTCTGTATT 1410

DB 509074 aaaggaatagtcgagtttagatgctcaatggttcaacacagacggaaggtctacaccc 509133

QY 1411 ATGATGCGCGGCAACGTTAAAGCAATACAGAGAGAGTACTTTAAGAGTCTTTC 1470

DB 509134 atgcacaccaggaacaaagctcaaaagcagatctgaagctatcagcttcaacaaacttgc 509193

QY 1471 ATTCTGTAGACTCTTTAGCGAGGGGTAAAGAAAGTTTAATTTGCTGCTTCTGACGAACT 1530

```

Db 509134 gttagctcttgccttgaggaagaaatagatgtgtccatctgaacagcgagagccaac 509253
Oy 1531 AAAATGTAGCCCTTATGTGTCGATCTCTTTGGATTAACCAAGGAATGCTTANGAA 1590
Db 509254 aaaaataaacttaacacctccctctgttctccaagaatagagagcaattttatgaa 509313
Oy 1591 AATACGAGCTTAGGAAAAACCAAGACTTTTCATTTTGTGCAAGCTCTGCTGGTACT 1650
Db 509314 agccatcacgaataaaccagcctcaacagccttctgtatctacgtcgtactgct 509373
Oy 1651 GCAACAACTACAGATGTCCAGCGTTTCCATAGACAGTCCATCCATATGGAT 1710
Db 509374 gctagcgatattatataatgaagcgcttctactctccagtaacaaactccagaactcat 509433
Oy 1711 CAAGTACTTGGGGAAGATCTGGGTGATGATACCCGACAGCTCCAAAGATTAAGACA 1770
Db 509434 tacgggtatccaggga---catlgygaagccactlgygcagacacatacactgcaaaatca 509490
Oy 1771 GCGACATAGCTTGGACCAATACAGGCTTTCGATCCGATCCGAGCGTCAGAGACTTTA 1830
Db 509491 ggaactaagactlgygtaacacagggctaacaccctaatcctgagcgtagagcttcgta 509550
Oy 1831 GTTCCATAATAGCTTTGGGATCTTTTTCAGACATCCAAAGCATTTCAAGTGTATAGAG 1890
Db 509551 gtcccgatctatataatgagcctcttacttacttgcacttgcactacagcaatcatgaca 509610
Oy 1891 AGAAGTCTTGTACTTGTGTTCAGATCGAGGCTTGGGCTGGCGGAGTCCCAATTTG 1950
Db 509611 tctcaagaaatagatlatcatcagcaagagactctggtacagagactcgcaatttc 509670
Oy 1951 TTAGTAAAGATTAAGAAAGGGGAAAAACGCAATACCTCATTAATCTGGTGGATATCT 2010
Db 509671 ttcataagaaataaactaagaaactaacaagactccgacataaagaactagactatct 509730
Oy 2011 ATCGAGGTCGACGCAAACTGTTCGAAACTTAATTAATTAATTAATTAATTAATTAAT 2070
Db 509731 gttagaggaagtgctgaagatcttcttgaataataatctcagtgtagcttctgcagctc 509790
Oy 2071 TTTGGTACGCTTAAGATTTCTTGTGCTGCTTAATAATCACTACTGATACCTAGCAGAGCC 2130
Db 509791 ttcgtaagaataaagactcttctatagltgaataaactctcaactatcttagcgcg 509850
Oy 2131 TTCTATATCCAAACATTTACAGATGTGTGCTTCAATAGTGTGCTTGTGATTAATCTT 2190
Db 509851 ctataaccgcaacatcgagccttccctaggaagactcccatgcccctcaatttggaaatctc 509910
Oy 2191 CCTGGCTCTTGGAGTCATTAACCCCTGCTTTAGAGGGGACGCTCGCTTATAGCCACGTC 2250
Db 509911 accgacatgctgaagaatatactccctcaatttgaatgcccgcagctaaagctaaagctaacct 509970
Oy 2251 AGTATGATCTGAAGACAAAGTATACGCTGATCTCTGAGTGAAGAGTTCTTGGGGGAAT 2310
Db 509971 aaaaatgatalgtagctcgctatacttccatctcctgaagctccttagcaacat 510030
Oy 2311 AATGCTTTTAAATGATGTGTGGAGCTCT---TCTCATTTTATCCGATTAATCTGAT 2367
Db 510031 aactctlgyggctcagagctcagagatctctgctcctatacttccctaaagaagacccg 510090
Oy 2368 TGTTTTATACCTTATGTCATATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2427
Db 510091 tctctcagagatattctccctctttaaagtccagagcgtctcagccgcaacaaac 510150
Oy 2428 TTTTCGGAGAAAGGTACAGAAAGATCTTTGATGAACAGCAACTCTTCAATTAATCT 2487
Db 510151 tttaaagagatggtcgctgagccgctttagatgtagagacctaagtaactgctct 510210
Oy 2488 TTGCTTATAGGGGTGAAGTGTGAGAGTCTCAGATGTATGTATGATCTTCTTATGATCTG 2547
Db 510211 atccctgctgagcattcgcttgaataaactcccgaaagatgaaaaaaatctcgagatt 510270
Oy 2548 ACTTATCCCTTGTGCTGATCTTATCCGCAATGATCCCAATGCTACATGACGACTTGTA 2607

```

```

Db 510271 tctctagcctaactggtgatgtatcgtatgtaaaatcccgcttgccgacttctctaatg 510330
Oy 2608 ATCAGCGAGCCTCTTGGGAAACTTATGCCAATTAATGACAGCAGAGCCTTGCAAGTG 2667
Db 510331 gtctagtgagcctcttgtagctcgtatgtaaaaccccgagacagagccttcttagca 510390
Oy 2668 CGTCAGGCACTACTACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2727
Db 510391 agtgcctggaagccatctgactctctccctcatgtagaactctctggaagagctgctat 510450
Oy 2728 GAAGTTCGTGATCCTCAAGGATTTATATGATGATGATGATGATGATGATGATGATGAT 2787
Db 510451 gactcgtggtctgacgacacatctcaaatgtagtltggtggtcgaagatactcatctag 510510
Oy 2788 GAGCCTCTCTCATGCTCA 2806
Db 510511 ttcctacttctccctaa 510529

RESULT 10
AAA30853
ID AAA30853 standard; DNA; 3000 BP.
XX
AC AAA30853;
XX
DT 29-AUG-2000 (first entry)
XX
DE Chlamydia antigen CPN100639 full length coding sequence.
XX
KW Chlamydia antigen; diagnosis; infection; community acquired pneumonia;
KW therapy; upper respiratory tract disease; bronchitis; sinusitis;
KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;
KW ds.
XX
OS Chlamydia pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 101..2893
FT /tag= a
FT /product= Chlamydia antigen CPN100639
XX
PN W0200032794-A2.
XX
PD 08-JUN-2000.
XX
PF 01-DEC-1999; 99WO-CA01147.
XX
PR 01-DEC-1998; 98US-0110339.
PR 01-DEC-1998; 98US-0110340.
PR 01-DEC-1998; 98US-0110427.
PR 01-DEC-1998; 98US-0110428.
PR 01-DEC-1998; 98US-0110438.
XX
PA (CONN-) CONNUGHT LAB LTD.
XX
PI Murdin AD, Oomen RP, Wang J;
XX
DR WPI; 2000-412339/35.
DR P-PSDB; AAY90240.
XX
PT Nucleic acids encoding polypeptide antigens from Chlamydia useful for
PT preventing, diagnosing and treating diseases such as community acquired
PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
PT asthma
XX
PS Claim 2; Fig 7; 174pp; English.
XX
CC This sequence encodes a Chlamydia antigen of the invention, designated
CC CPN100639. The nucleic acids (and their complementary sequences) may be
CC used as diagnostic agents for detecting the presence of nucleic acids
CC encoding Chlamydia antigens in samples according to standard methods,
CC and therefore, for diagnosing Chlamydia infections. For example, they may
CC be used as primers and probes for diagnostic polymerase chain reaction

```





OY	178	ACGGAGATATTA	CTCCGCAAA	ACCTTGGGATTC	GGCAGCTTAA	CGAAGGCTGT	TTT	237
OY	178							
Db	181	tcaggaaatg	ctataataa	cgaatcgtc	ggaaagaa	cagcagctaa	cagctgcgtc	240
OY	238	TCGTGACAT	TACGGAATCT	TTTAAGCTT	TCCGGTAAG	GGGATACAT	CTTCTTTTAAAT	297
Db	241	acaggaacta	agggtg	atctgcga	atcttaac	ctggaaagg	atacctatc	300
OY	298	ATTAA	GCTAGTGC	---	TGAAGCG	CAGCACTT	TCTGTATCA	354
Db	301	gtatgtc	ggttcga	atgcga	tcgagc	tcgagcga	caacaatc	360
OY	355	CTAAGAGAT	TTTTGAG	CTTACTT	TTTAGCGG	CCCCATAC	GGGTATAC	414
Db	361	ttcaagaat	atcttctta	acattc	cttctta	ctgcagctc	-----	414
OY	415	TCAGAAAG	AGTGCAG	TAAATG	TGAGGGAG	ACTTACAT	TTGTATTA	474
Db	415	tcaggaaaa	agtaact	taagtat	ctgcagag	ccttaaat	cttaacgata	474
OY	475	TTATTAA	CAAGATT	ACTGTG	AGAA-----	AATGGCG	AGCCATTT	525
Db	475	ctctttag	caaaa	gcgtcc	caatga	agatca	taacatg	534
OY	526	AATCTTT	TTTAAAA	ACAGC	ACGCGAT	TCATTTCT	TTTGAAG	585
Db	535	actcttctat	ctcttgcg	gaatac	ctcttctata	accttca	ctagatag	591
OY	586	ACAGGAA	AAAAAG	TGGGGCT	ATTGTG	TGAC	GTAGATAT	645
Db	592	-----	aaatga	gttga	gcgact	ctgcgag	ctgcga	645
OY	646	GCTCCT	ACCCTCT	TCGAC	AAATATT	GTGAAG	CTGAC	705
Db	646	ggcagat	gactctta	tgataa	taaaag	gaac	tcgggg	705
OY	706	GGAAC	CTACAA	TTTAC	AGGAA	TACGTCT	CTGTATTT	765
Db	706	gccagct	ctccctga	cttaac	caaaa	taagctc	cttctt	765
OY	766	ACCCG	AGAAAT	TGAGAG	AGCTTT	CTTGAG	AG-----	810
Db	766	gctgcag	gcgaagg	gcggg	ccatttat	ctg	gaaaaa	825
OY	811	ATTAT	CGGAA	TCAGAG	TGTAAC	TTTCTC	AGAAAC	870
Db	826	atctctg	aaataa	agctgc	agcttc	gcgag	aactct	885
OY	871	ATTAT	AGCTAA	GAAC	CTTTAC	CTG	GGCTCC	930
Db	886	atctgtg	cccatg	ctctaga	ctcttcc	gcgtc	gtgcgc-----	939
OY	931	AATAT	ATGCCA	AGGTACC	ACTCG	AGTAAT	GTGAG	990
Db	940	aatagat	gtcgg	aaaca	agctgtc	gagga	gcgaag	999
OY	991	GAGGT	AGCTTT	TACGAG	AGAGG	AGGAA	TATAC	1050
Db	1000	tcttaag	ctctctc	tg	aaataa	agagata	caagctc	1055
OY	1051	ACTAC	ACCAC	AACTAC	AAAAA	AAATTT	CTATG	1110
Db	1060	acctccg	cgcaaca	tcgaca	cggaat	gtctat	ctactc	1119
OY	1111	AATTT	AGTGC	AAATAT	TGGGCA	TATG	CACTTT	1170
Db	1120	aactta	agggcag	cccaag	cccaat	ctatctat	ctatgc	1179
OY	1171	GCTCG	GGATTCT	CAAG	TA	CTTTAA	TTTCA	1230
Db	1180	acagga	gag-----	cttcag	gcgtctc	gcacat	caaca	1236
OY	1231	TATAG	TGGTGC	ATTGTTT	TTTCTG	TGA	AAAGCTCT	1290

[illegible]





OY	526	AACTCTTTGGAAAAACAGACAGCGGATTCGATTTCTTTTGAAGGGAATAATTCAGCGCA	585
OY	526		
Db	535	acctcttcatcttcyggaaatacctctctcataacctcaactagtaatagcgcacaa---	591
OY	586	ACAGGAAAAAGGTGGGGCTATTGTGCTACTGACTGTAGATTTTACATATATACG	645
Db	586		
Db	592	-----aaattagtgtagcgatctatagctctgcgcgtaagatttccagaaacacc	645
OY	646	GCTCTTACCTCTCTCGAACCAATATGTGCGAAGCTGCGAGTGGAGACTTAAATAGCACA	705
Db	646		
OY	706	GGAAACTGTACAAATTTACAGGGGAATACGCTCTCTGTATTTTTCGAAAAATAGTGTGACAGG	765
Db	706	gcccagcttcgattactactcaaatagctcccttctctcttccttccttcygaacacgtcgaacgagt	765
OY	766	ACCCGAGGAATTGAGAGAGCTTTCTCGAGA-----TCCCATGTTTAC	810
Db	766		
OY	811	ATATCTGGGAATCTAGAGTGTAACTTTCTCGAAGAACCAAGCTGTACTTAATGGCGAGCC	870
Db	826	atctctggaataaagctctgcaccttcgcgcgagaactcttcagtaactaaagcggagca	885
OY	871	ATTTATGCTAAGAACTTTACACTGTGGCTTCCGGGGGGGGGGATCTCTCTTTCTTAC	930
Db	886	atctgctgccatgctctagatctcttcgcgtcgtcgccc-----tacctatttcaaat	939
OY	931	ATATPATGCCAAGGTACCACGCGAGGAATAGTGAGGACATTTCTATACGTGCACACTGCA	990
Db	940	aatagatgcygggaacacagctgcgcggaagggcgcgcatcttgcaattgcgcactctgga	999
OY	991	GAGTGTACTCTTTAGCGAAGAGCGAGGGGACATTACCTTAAATGGGAATCCATTTCTTCA	1050
Db	1000	tcttaagctctctcgtcaaatcaagsgagacatcagcttcccttgccaacactcaactca	1059
OY	1051	ACTTACACCAAACTACAAAAAAGAAATCTTATGTACATAGATCTTACTCCAAAAGTCAAG	1110
Db	1060	acctccgcgcacacatcgacaacggaatagctatctacctgggagtcgcgcagcaaaataacg	1119
OY	1111	AATTTACGTCAAAATATCTGGGCATATGACATCTTTTACAGATCCGATTTACTCTATACG	1170
Db	1120	aacttaagsgagcccaagccaaactcatctatctattctatagctcggttcacatcaacc	1179
OY	1171	GCTCGGATTTTACAGATCTTTTAAATCTCAATAGAGCTGATGACAGTAAATGTCACAGT	1233
Db	1180	acagagag---cttaagacgctctgcaccatcaacaaacggaatagaaactcgcgtttagt	1236
OY	1231	TATGTGGGTGATTTGTTTCTTGCGAAAACTCTCGAAGATGAAACAAATTTGCA	1290
Db	1237	tattcaggaacgatgtattcttcttcyggaaagctctctgcgcagatggaagcgaagcgtct	1296
OY	1291	GACAACCTCACTTACGCTGAGAGCAACGCTGTAACTCTTAACTGACGAATTTAATCACTT	1350
Db	1297	gatacttcacatctatataaagaacacatcgtctctagctctcgcgaaccttagcctac	1356
OY	1351	AAAGTGTGTGCACTCTCGCATAGGAAGGCTTTTACTCACACCGCGGGTCTCTCTTATTT	1410
Db	1357	aaaggaatacgtcgaattagatgtcaatggtttcaacagacctgaaagcctctacactctc	1416
OY	1411	ATGATCTCGGGCAACAGCTTAAAGCAAGTACAGAGAGAGTCTACTTTAAACAGTCTTTC	1470
Db	1417	atgcaaccaggaacaaagcctcaagaagatcgaagcctatagctcttaaccaaatgttc	1476
OY	1471	ATTCCTGTAGACTCTTAGCGGAGGGTAAAGAAATGTATATGTGCTCTTGACAGCAAGT	1530
Db	1477	gttcatcttctcgtcttagagggaaataagagtggttccattgaaacagcagagccaac	1536
OY	1531	AAAAATGTAGCCCTTATGTGTGCTCGGATTTCTTTTGGAAATACCAAGGGAATGCTATGAA	1590
Db	1537	aaaactataactcaacctctctcttgcttcttcccaagaatagtcggaattttatgaa	1596
OY	1591	AATCAGCACTTAGAAAAACATCAAGACTTTTCAATTTGTGACACTCTCTGCTGGTACT	1650

Db	1597	agcca -tacgtaaaccaagcccttcacgcaagcccttcggtgtatctacitcgtcactcgtc	1655
Qy	1651	GCACACACTACAGATGTTCCAGCGGTTCTACAGTAGCAACTCTCAGCATATGGGAT	1710
Db	1656	gtacgcgatattatcatcogatcgctcttcacctctccagtaacaaactccagaacctcat	1715
Qy	1711	CAAGGTACTTGGGAATGACTTGGGTTGATGATACCGGAACGACCTCCAAAGACTTAAGCA	1770
Db	1716	tacggtatccaggga---catltyggaagccactltygcagacatcaatcaatcaatca	1772
Qy	1771	GGACATTAGCTTGGACCAATACAGCGCTACCTTCGAAATCTGAGGCTGACAGACTTTA	1830
Db	1773	ggaactgatcttgcgttaactcaggggtacaacaccttaactcaggtagagcttccgta	1832
Qy	1831	GTTCTATATACCTTTTGGGATCTTTTTCAGAAATCCAAACCGATTCCAGGTCTCATACG	1890
Db	1833	gtccgcattcaatltaatlyggatccttcaatcgcacttcgaactcaagaagatcagaa	1892
Qy	1891	AGAAAGCTTTGACTCTTTTGTTCAGATCGAGGCTTGGGGCTGCGGAGTCCCAATTC	1950
Db	1893	tctcaaggaatagtatcatcatcagcaagagagctcttggtacatcagaaatcgcgaattc	1952
Qy	1951	TTAGATAAAGATAAAGAAAGGGAAAAACGCAATACCGTCAATTAATCTGCTGATATCT	2010
Db	1953	ttcccaatggaataatcaaggaaactaaaccaagcatctccgacataaaagctacgtcata	2012
Qy	2011	ATCGAGGTGCGAGCGCAACCTGTTTCGAAATCTTAATTAAGTTTGCCCTTTGGCCACTC	2070
Db	2013	gttgaaggaagtgctgaagatcttctcgaataatctcaagtgtagctcttcgcagctc	2072
Qy	2071	TTTGGTACCGATAAAGATTTTCTTAGTCGTAAAAATCATACTGATACCTATCGACGAGCC	2130
Db	2073	ttcgttaagaaagtaagccgtttatagttaagaaataactctcaataactatttagcgtc	2132
Qy	2131	TTCTATATCCACACATTTACAGATGATGAGGGCTCATAGTTGTCTCTTGATATAACTT	2190
Db	2133	ctataaccggaacctcgagatctccttagagagcttcccaatccctcaatcttlygaatlc	2192
Qy	2191	CGTGACTTGGAGTCTAATTAACCCCTGTGTTGAAGGGCAGCTCGTTATAGCCAGTC	2250
Db	2193	accgacatcgtgaagatcttccctctcatlittgaatlygcacagctaaagctacacact	2252
Qy	2251	AGTATGATCTGAAGCAAAAGTATCTACGCTATACCTCGAGAGTGAAGGTTCTGGGGGAA	2310
Db	2253	aaaaatgatataagatactcgtactcctcaatcctgaagctcaagagctcttgcagccat	2312
Qy	2311	AATGCTTTAATCAATGATGTTGGAGGAGCTCT--TCTCATTTTATTCGTAATACCTGAT	2367
Db	2313	aactctggtgcttagagctcgcgaagatctcgtcctatctatctcccaagaagcaccg	2372
Qy	2368	TGTTTTGATACCTATGCTCCATATCAATCAATGATATCTACTATATATAGTCAGACAGC	2427
Db	2373	ttcttccagggatattccctctcttaagtttccagagcagttctaaagccgccaacaaac	2432
Qy	2428	TTTCGGAGAAAGGTACAGAAAGAAAGTCTTTTGATGACAGACACTCTTCAATTATCT	2487
Db	2433	tttaagaagaaatgagctcgaagccgctgccttgaagagagagactagtgaaatcgtct	2492
Qy	2488	TTTGCTATAGGGGGTGAAGTTTGAAGTTCTCTGATTTGTAATGACTTTTCTATGATCTG	2547
Db	2493	atccctgtcgcacatcgtgttagaanaaaatctccgaagatgtaaaaaataatttcgaagt	2552
Qy	2548	ACTTTATCTATGTCTCTGATCTTATTCGCAATGATCCCAATGCACTACAGACT--TG	2605
Db	2553	ttctatacctacatctgtgtatgtatcgttlaaaaaatccccgttcgcgtactcttaaty	2612
Qy	2606	TAAATCAGGAGGAGCTCTTGGGAACCTTAATGCCAATTAATCTTACACGACAGCGCTTGAG	2665
Db	2613	gtctcaagtgaagcctcttgcacttgcattagtaaaaaactcgcagaaagccttcttag	2672
Qy	2666	TGCGTGAGGCGAGTCACTACGCTTCTCTCATATGTTGAAGTGCCTCGGCAAGTTTGCT	2725









Db	1566	cacttattggaccggacggcagcgtttatagtaaatcatagtttaagaatccctcagtc	1625
Qy	1617	CTTTTCACTTTTGACAGCTCTCTGCTGGGTAAGTCGACAACTACAGATGTTCCAGCGGT	1676
Db	1626	ctacgacacttaagactccaagcttcttgaaactctgaacagcaacgcgcagtgactccaga	1685
Qy	1677	TCTCACTATGACCACTCTCTACGCACTATGGGTATTCAGGTACTTGGGATACTTGGGT	1736
Db	1686	tcctataatgggttagaagaattccattacggtctatcaggaacttggggcccaattgttgg	1745
Qy	1737	TGATGATACCGCAAGCACTCCAAAGACTAAGACAGCGACATTAGCTTGGACCAATACAG	1796
Db	1746	gg-----ggaacagggtctcttaccgactgtaacacttcaactggaactgaataactcgg	1793
Qy	1797	CTACCTTCCGAATCCTGAGGTCGAAGACCTTAGTTCCTAATAGCTTTGGGAGATCTTT	1856
Db	1794	ctatacttcatactccgaagcgtatcgctctttagtccctaatagcttatgtaatgcctc	1853
Qy	1857	TTCAACATCCCAAGCATTCAGGTGTCATAGAGAGAAGTGTTCATCTTGTTCAGA	1916
Db	1854	tatagataatagctctctccattactctatagtagactgcaacggaaggttgcaggysga	1913
Qy	1917	TCGAGGCTCTGGGCTCGGGAGTCCGCACATTTCTTAGTAAAGATTAAGAAAGGGAAA	1976
Db	1914	ccggtcttcttgggttgcgtgatalactaactctccataagatagatctacaaaacgcg	1973
Qy	1977	ACCGAAATACCGTCAATTCGTGTGGATATGCTATCGAGAGTGCAGCGCAAACTGTTC	2036
Db	1974	acgggggttcgcacattgtagtggcgttatgcatagaaggaaccatactactgtctc	2033
Qy	2037	TGAAACTTAATTAAGCTTGTGCTTTTGGCCAACTCTTGTGATCGATTAAGATTTCTTAGT	2096
Db	2034	agataagattcttagtgcctcattcttgcaagctctcttggaaagataagactacttgtt	2093
Qy	2097	CGCTAAAAATCATCTGATATCATGACGAGGCTCTATATCCAAACATTAACGAATG	2156
Db	2094	agctaaagaataaggtacagctcagcaggaagaaactctataccscgacaaacgaacta	2153
Qy	2157	TAGTGGGTCATAGTGTGTCCTTAGATAAA--CTTCCGTGGCTTGTGAGTCATATAAC	2213
Db	2154	tatctctcttcccttgcaacaactacgacctgttgctgtcttactgtcttccataagagatcc	2213
Qy	2214	CTGTGTTTAGAAGGAGAGCTGCGCTTATAGCCACGTCATATATGATTCGAACAAATA	2273
Db	2214	tgtctctcttccaggaaactttagcaaccatacagatacagtaacgtctgaaacccaagta	2273
Qy	2274	TACTGCGTATCTGAGGTGGAAGTTCTTGGGGGAATATGCTTTTAACATGATGTGGG	2333
Db	2274	tacaacatctccactcgtttaaaggaagcgtggggaatgtagttctgcgtttagaattcgg	2333
Qy	2334	AGCTTCTTCATTTCTTATATCCGAATACCTGCATGTGTTGATATACCTTAGTCCATACAT	2393
Db	2334	tgyaagaagcttcgatttgccttagatgaaagtgctcatatttgacagtaacagccctcat	2393
Qy	2394	CAAACTGAATCTGACCTAATATACGTAGAGCAAGCTTCTGCGAAGAAAGGTACAGAAAGAG	2453
Db	2394	gaaatctcagtttgcctatgcaatacaaggaagttttaaagaacaaggaacgaagctcg	2453
Qy	2454	ATCTTTTGATGACAGCAACCTCTTCAATTATCTTGCCTATAGGGGTGAAGTTTGAGA	2513
Db	2454	tgaatttgyaagtagcgcgtcttgatccttgcttaccatacgtacgtacgtattgttaa	2513
Qy	2514	GTTTCTGATTTGAATGACCTTTCTTATATATCTGACTTATACCTTAGTCTGATCTTAT	2573
Db	2514	ggaactgacactgcgaagatgcaacgtacaaactaactcttggtatactgtagctctgt	2573
Qy	2574	CCGAAATGATCCCAAAATGCACTACAGCACTTGTATATCAAGCGAGCCTTGTGGAAACTTA	2633
Db	2574	tcgtagtaaacccgactgttagcacaacactcgtgaattgcggtatgtcttgyaanaactc	2633
Qy	2634	TGCGAATTAATTAGCAACGAGGCTTGCAAGTGGGTGAGGCAAGTACTACGTACGCTTCTC	2693

Db	2634	cggtacggaattgggaagaacaagcttagtccttcgcgcgaagggaacattttgcttaa	2693		
Oy	2694	TCCTATGCTTGAAGCGTCGGCGCAAGTGTGCTTTGAAGTTCGTGACACCGAGGATT	2753		
Db	2694	ctcaattttgaagaccttagcacaatttcttcttgattgcgctgggactcgcacatta	2753		
Oy	2754	TAAATAGATCTTGGGGGTAAAGTTCCAATCTTA	2786		
Db	2754	caatgtagacttagagacaataaccattctcta	2786		
RESULT 15					
AA	27021	AAA27021 standard; DNA; 3000 BP.			
AC	AAA27021;				
XX	11-AUG-2000	(first entry)			
DE	Chlamydia pneumoniae 98kD putative outer membrane protein gene.				
XX	Chlamydia; antigen; vaccine; infection; outer membrane protein; ds.				
OS	Chlamydia pneumoniae.				
XX					
XX	Key	Location/Qualifiers			
FT	CDS	101..2887			
FT		/*tag= a			
FT		/product= "98kDa putative outer membrane protein"			
FN	WO200026237-A2.				
PD	11-MAY-2000.				
XX	29-OCT-1999;	99WO-GB03579.			
XX	29-OCT-1998;	98US-0106070.			
PR	01-MAR-1999;	99US-0122066.			
PR	27-OCT-1999;	99US-0428122.			
PA	(CONN-) CONNAUGHT LAB LTD.				
PI	Murdin AD, Oomen RP, Dunn PL;				
DR	WPI: 2000-365569/31.				
DR	P-PSDB; AAY94327.				
XX	Novel Chlamydia 98 kDa putative outer membrane protein antigen, used				
PT	for vaccination and protection against Chlamydia infection				
PS	Claim 1; Fig 1; 93pp; English.				
XX	The present sequence is the 98kDa putative outer membrane protein gene				
CC	from Chlamydia pneumoniae. The genomic sequence was amplified using two				
CC	PCR primers. The 5' primer contains a NotI restriction site, a ribosome				
CC	binding site, an initiation codon and a sequence close to the 5' end of				
CC	the 98kDa putative outer membrane protein coding sequence. The 3' primer				
CC	contains the sequence encoding the C-terminal sequence of the putative				
CC	outer membrane protein and a BstXI restriction site. The stop codon was				
CC	excluded and an additional nucleotide was inserted to obtain an in-frame				
CC	C-terminal fusion with the Histidine tag. The PCR product was cloned				
CC	into a eukaryotic expression vector (pCA-Myc-His) by restricting both				
CC	the vector and the PCR product with NotI and BamHI and performing a				
CC	ligation reaction. This expression vector was injected intramuscularly				
CC	and intranasally into mice, which were subsequently inoculated with				
CC	Chlamydia pneumoniae. The chlamydial lung titers of the immunised mice				
CC	were lower than those of the controls. Thus the 98kDa putative outer				
CC	membrane protein can be used as a vaccine to provide protection against				
CC	Chlamydia infections, especially Chlamydia pneumoniae infections.				
CC	The polypeptide may also be administered orally to treat Chlamydia				
CC	infection. The present sequence may also be used in the				
CC	construction of attenuated Chlamydia strains that can over-express the				
CC	gene or express it in a non-toxic form.				





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 03:25:12 ; Search time 268.88 Seconds  
(without alignments)  
1981.965 Million cell updates/sec

Title: US-09-446-677B-3

Perfect score: 2815  
Sequence: 1 ATGAATGCGCAATTTTCCTG.....CTCATCTCAGCAATTTCTG 2815

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40.8	1.4	1817	2	US-08-743-637B-1
2	40.8	1.4	1817	3	US-08-526-840B-1
3	37.2	1.3	7218	1	US-08-232-463-14
4	35.8	1.3	1365	2	US-08-472-172-3
5	35.2	1.3	7833	1	US-08-074-879-9
6	35.2	1.3	7833	1	US-08-468-057A-9
7	34.2	1.2	1347	1	US-08-472-172-5
8	34.2	1.2	3771	1	US-07-876-280-5
9	34.2	1.2	3771	1	US-08-049-783-1
10	34.2	1.2	3771	1	US-08-158-232-5
11	34.2	1.2	3771	1	US-08-304-626-5
12	34.2	1.2	3771	1	US-08-316-301A-5
13	34.2	1.2	3771	1	US-08-611-928-5
14	34.2	1.2	3771	4	US-09-173-891-5
15	34.2	1.2	3771	4	US-09-076-137-5
16	34.2	1.2	3771	5	PCT-US92-03624-5
17	33.6	1.2	1125	3	US-08-591-685-10
18	33.6	1.2	2157	1	US-08-451-715A-3
19	33.6	1.2	7218	1	US-08-232-463-14
20	33.4	1.2	1884	1	US-08-257-073-8
21	33.4	1.2	5097	4	US-09-315-793-11
22	33.2	1.2	381	1	US-08-500-222-7
23	33.2	1.2	381	1	US-08-500-125-7
24	33.2	1.2	381	2	US-07-779-704B-7
25	33.2	1.2	747	1	US-08-500-222-3
26	33.2	1.2	747	1	US-08-500-125-3
27	33.2	1.2	747	2	US-07-779-704B-3

28	33.2	1.2	1008	4	US-08-235-836C-129	Sequence 129, App
29	33.2	1.2	1131	1	US-08-500-222-5	Sequence 5, Appl1
30	33.2	1.2	1131	1	US-08-500-125-5	Sequence 5, Appl1
31	33.2	1.2	1131	2	US-07-779-704B-5	Sequence 5, Appl1
32	33.2	1.2	1497	1	US-08-500-222-1	Sequence 1, Appl1
33	33.2	1.2	1497	1	US-08-500-125-1	Sequence 1, Appl1
34	33.2	1.2	1497	2	US-07-779-704B-1	Sequence 1, Appl1
35	33.3	1.2	2380	1	US-08-572-951-3	Sequence 3, Appl1
36	32.8	1.2	2617	1	US-08-430-024-1	Sequence 1, Appl1
37	32.8	1.2	2617	1	US-08-782-009-1	Sequence 1, Appl1
38	32.8	1.2	2617	3	US-09-017-302-1	Sequence 1, Appl1
39	32.8	1.2	3730	1	US-08-242-932-8	Sequence 8, Appl1
40	32.8	1.2	3730	1	US-08-714-481-8	Sequence 8, Appl1
41	32.8	1.2	3730	5	PCT-US95-06111-8	Sequence 8, Appl1
42	32.8	1.2	4200	1	US-08-242-932-1	Sequence 1, Appl1
43	32.8	1.2	4200	1	US-08-714-481-1	Sequence 1, Appl1
44	32.8	1.2	4200	5	PCT-US95-06111-1	Sequence 1, Appl1
45	32.6	1.2	678	2	US-08-951-871-3	Sequence 3, Appl1

#### ALIGNMENTS

```
RESULT 1
US-08-743-637B-1
; Sequence 1, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OUELLETTE & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586, 90012
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1817 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecalis
; US-08-743-637B-1
```



Query Match	1.4%	Score 40.8;	DB 2,	Length 1817;
Best Local Similarity	51.1%;	Pred. No. 0.012;	Mismatches 92;	Indels 0;
Matches 96;	Conservative 0;	Mismatches 92;	Indels 0;	Gaps 0;
QY 1405	GTTATTATGATGCGGACACAAGTAAAAAGCAAGTACAGAGAGGTCACCTTAAACAGT	1464		
Db 545	GATAAAAACGGCGAGTAAACAAGTGTACAGAAATCCAGTTGACCAAGTTGTGAAAATT	604		
QY 1465	CTTCCATTCCCTGTAGACTCTTTAGGCGAGGTAAAGATGTAATGCTCTCTGCA	1524		
Db 605	CAATCGTTGATGCAATGCGTGAAAGAGATTAATAAAATATTGCTCTCTGATTAATCA	664		
QY 1525	GCAAGTAAAAATGTAGCCCTTAGTGTCGATTTCTCTTTGGATTAACCAAGGAATGCT	1584		
Db 665	GAAACTAAAGATCTGTCTTTTACTATTGCAACAAGGTGTAATTAATGAAGGCAATTA	724		
QY 1585	TATGAAAA 1592			
Db 725	TTTTTATA 732			

```

: RESULT 2
: US-08-526-840B-1
: Sequence 1, Application US/08526840B
: Patent No. 6001564
: GENERAL INFORMATION:
: APPLICANT: BERGERON, Michel G.
: APPLICANT: OUELLETTE, Marc
: APPLICANT: ROY, Paul H.
: TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
: TITLE OF INVENTION: APPLICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
: TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
: TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
: NUMBER OF SEQUENCES: 177
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: QUARLES & BRADY
: STREET: 411 East Wisconsin Avenue
: CITY: Milwaukee
: STATE: Wisconsin
: COUNTRY: USA
: ZIP: 53202-4497
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/526,840B
: FILING DATE: 11-SEP-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/304,732
: FILING DATE: 12-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: BAKER, Jean C.
: REGISTRATION NUMBER: 35,433
: REFERENCE/DOCKET NUMBER: 850586.90012
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (414) 277-5000
: TELEFAX: (414) 277-5591
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1817 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Enterococcus faecalis
: US-08-526-840B-1

```

		Query Match	Similarity	1.4%;	Score	40.8;	DB	3;	Length	1817;	
		Best Local	Similarity	51.1%;	Pred.	No.	0.012;				
		Matches	96;	Conservative	0;	Mismatches	92;	Indels	0;	Gaps	0;
OY	1405	GTTATTATGATGGCGGCACAAGCTTAAAGCAAGTACAGAGAGGTCACTTAAACAGGT	1464								
Dd	545	GATATAAAGCGGGTCAGTAAACAAGTGTTACAGGAATCCAGTGATCCAACTTGAAAAATT	604								
OY	1465	CTTTCATTTCCGTGTGACACTTTTAGCGCGGGTAAGAAGTTGTATTGCTGCTCTGTGCA	1524								
Dd	605	CAATCGGTTGATGCATCGGTGCAAGAGAGAGTTAAAANAATTTGTTGCTTCTATATATCCA	664								
OY	1525	GCAGTAAAAAATGTAGCCCTTAGTGTGCCGATTCCTTTTGATPAACCAGGAATGCT	1584								
Dd	665	GAAACTAAAAAGATCCTTGCTTTTACCTATTGACAAACAGGTAAAAATAGCAAGCGCAATTA	724								
OY	1585	TATGAAAA	1592								
Dd	725	TTTTATAA	732								

```

1      RESULT 3
2      US-08-232-463-14/c
3      : Sequence 14, Application US/08232463
4      : Patent No. 5670367
5      : GENERAL INFORMATION:
6      :   APPLICANT: DORNER, F.
7      :   APPLICANT: SCHEIFLINGER, F.
8      :   APPLICANT: FALKNER, F. G.
9      :   TITLE OF INVENTION: RECOMBINANT FOMLPOX VIRUS
10     :   NUMBER OF SEQUENCES: 52
11     :   CORRESPONDENCE ADDRESS:
12     :   ADDRESSEE: Foley & Lardner
13     :   STREET: 1800 Diagonal Road, Suite 500
14     :   CITY: Alexandria
15     :   STATE: VA
16     :   COUNTRY: USA
17     :   ZIP: 22313-0299
18     :   COMPUTER READABLE FORM:
19     :   MEDIUM TYPE: Floppy disk
20     :   COMPUTER: IBM PC compatible
21     :   OPERATING SYSTEM: PC-DOS/MS-DOS
22     :   SOFTWARE: PatentIn Release #1.0, Version #1.25
23     :   CURRENT APPLICATION DATA:
24     :   APPLICATION NUMBER: US/08/232,463
25     :   FILING DATE:
26     :   CLASSIFICATION: 435
27     :   PRIOR APPLICATION DATA:
28     :   APPLICATION NUMBER: US/07/935,313
29     :   FILING DATE:
30     :   APPLICATION NUMBER: EP 91 114 300.6
31     :   FILING DATE: 26-AUG-1991
32     :   ATTORNEY/AGENT INFORMATION:
33     :   NAME: BENT, Stephen A.
34     :   REGISTRATION NUMBER: 29,768
35     :   REFERENCE/DOCKET NUMBER: 30472/114 IMMU
36     :   TELECOMMUNICATION INFORMATION:
37     :   TELEPHONE: (703)836-9300
38     :   TELEFAX: (703)683-4109
39     :   TELEX: 899149
40     :   INFORMATION FOR SEQ ID NO: 14:
41     :   SEQUENCE CHARACTERISTICS:
42     :     LENGTH: 7218 base pairs
43     :     TYPE: nucleic acid
44     :     STRANDEDNESS: single
45     :     TOPOLOGY: linear
46     :   IMMEDIATE SOURCE:
47     :     CLONE: pTZ9pt-FLS
48     :   US-08-232-463-14

```

	Best Local Similarity	0.8%; Pred No.	0.4;	Matches	3; Conservative	213; Mismatches	156; Indels	0; Gaps	0.
Oy	416	CAGGAAAAGTCCACTTAATGTGAGGGGATCTTACATTGTGAACAATGGAATAATT	475	:	:	:	:	:	:
Db	1435	CGRRR	1376	:	:	:	:	:	:
Oy	476	TATTTAAACAGATTACTGTGTGAGAAAATGGGGGCCATTTCACCAAGAATCTTCTT	535	:	:	:	:	:	:
Db	1375	RRR	1316	:	:	:	:	:	:
Oy	536	TGAAAAAAGCAGCGGATGCATTTCTTTGAAGGAATTAATCAGCGCAACAGGAAAA	595	:	:	:	:	:	:
Db	1315	RRR	1256	:	:	:	:	:	:
Oy	596	AAGTGGGCTATTGTGTACTACTGTACTGTATTACAATAATACGGCTCACCC	655	:	:	:	:	:	:
Db	1255	RRR	1196	:	:	:	:	:	:
Oy	656	TCTTCTCGAACAAATWTGCTGAAGCTGACGTGAGTATAAATAGCAGCAAGAACTGTA	715	:	:	:	:	:	:
Db	1195	RRR	1136	:	:	:	:	:	:
Oy	716	CAATTACAGGGAATCGCTCTTGTATTTTTCGAAAATAGTGTGACAGCAGCAGGAA	775	:	:	:	:	:	:
Db	1135	RRR	1076	:	:	:	:	:	:
Oy	776	ATGGAGAGCTC 787		:	:	:	:	:	:
Db	1075	RRRRRRRRATC 1064		:	:	:	:	:	:

RESULT 4  
 US-08-472-172-3  
 Sequence 3, Application US/08472172  
 Patent No. 5985288  
 GENERAL INFORMATION:  
 APPLICANT: Munson, Jr., Robert S  
 APPLICANT: Grass, Susan  
 APPLICANT: Chong, Pele Y Y  
 APPLICANT: Fahm, Raafat  
 APPLICANT: Sia, Charles D Y  
 APPLICANT: McVerry, Patrick  
 TITLE OF INVENTION: Kleih, Michel  
 TITLE OF INVENTION: Outer Membrane Protein P1 and Peptidases  
 of Haemophilus influenzae Type B  
 NUMBER OF SEQUENCES: 26  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Slim & McBurney  
 STREET: Suite 701, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R6  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/472.172  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/849,411  
 FILING DATE: 07-JUL-1992  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Stewart, Michael I  
 REGISTRATION NUMBER: 24973  
 REFERENCE/DOCKET NUMBER: 1038-471  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (416) 595-1155

```

? TELEFAX: (416) 595-1163
? TELEX: 063-24667 SIMBAS
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1365 base pairs
? type: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
US-08-472-172-3

```

Query Match	1.3%	Score 35.8	DB 2	Length 1365
Best Local Similarity	47.2%	Pred. NO. 0.38		
Matches 109	Conservative	0	Mismatches 122	Indels 0
			Gaps	0
QY 1018	GACATTACCTTCATGGGAATGCCATTGTGGCACTACACCAACAACATCAAAAAAGAAAT			1077
Db 223	GATGTAGAAATTAATATGATGATGTAACTGCTTATAGCAACTCAATCAATGAAC			282
QY 1078	TCTATTGCATATAGATCTACTGCAAAAGATCAGCAATTATACGTGCAATATCTGGCGATAC			1137
Db 283	TCAGCAAGTACGGCGTCAGCTTCAAGGCGTAAATGTTGTCGCGCTTTTGTGCCAAAT			342
QY 1138	ATCTTTTCTACGATCGCATTTACTGCTAATACGGCTGGGATTTACAGATCTTAAAA			1197
Db 343	CTTTATTTTCGTTGCTCCACGTAATATATAATTCGGGCTGGGCGCAGGAATGATGTAAAT			402
QY 1198	CTCAATTAAGGCTGATGCAAGTATATGTCACAGTTTATAGTGGTCCATTGTT			1248
Db 403	TTTCGCTCAAAAAAGTAATATACATAGTTAATGATGCTGGTATATTTGGT			453

RESULT 5  
 US-08-074-879-9/c  
 Sequence 9, Application US/08074879  
 Patent No. 5656423  
 GENERAL INFORMATION:  
 APPLICANT: Orth, Gerard  
 APPLICANT: Volpers, Christoph  
 APPLICANT: Streek, Rolf  
 TITLE OF INVENTION: DNA Sequences Derived from the Genome of  
 TITLE OF INVENTION: the Papillomavirus HPV39, Their Use in In Vitro Diagnosis  
 TITLE OF INVENTION: and for the Production of an Immunogenic Composition  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 ADDRESSEE: Dunner  
 STREET: 1300 I Street, N.W.  
 CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20005-3315  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/074,879  
 FILING DATE: 16-JUN-1993  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO 92/1136  
 FILING DATE: 09-JUL-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Potler, Jane E.R.  
 REGISTRATION NUMBER: 33,332  
 REFERENCE/DOCKET NUMBER: 02356, 0066-00000  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-408-4000  
 TELEFAX: 202-408-4400  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:

```

; LENGTH: 7833 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-074-879-9

```

Query Match	1.3%	Score 35.2	DB 1	Length 7833
Best Local Similarity	58.7%	Pred. No. 1.8		
Matches 61	Conservative 0	Mismatches 43	Indels 0	Gaps 0

OY	333	TACACATGATAAAAAATCTGTGGCTAACAGAAATTTCGACTTCTTAATTTTGCGGCCCC	392
Db	5780	TCCACTGTTAATATATCTAAGCTGGCAGCATATATATATGCTGCCTGCTCGTTAACATA	5722
OY	393	ATCATCGGTATACACAACCCCCCTTCAGAGAAAAGTGCATTAAT	436
Db	5720	ATCATCAAGTATTGACAACCTTCCCAACAGAAAGTGGAGGCAAT	5677

RESULT 6  
US-08-468-057A-9/c

Sequence 9, Application US/0846805/A  
Patent No. 5665535  
GENERAL INFORMATION:  
APPLICANT: Orth, Gerard  
APPLICANT: Volpers, Christoph  
APPLICANT: Streeck, Rolf  
TITLE OF INVENTION: DNA Sequences Derived from the Genome of  
the Papillomavirus HPV33, Their Use in In Vitro Diagnosis  
TITLE OF INVENTION: and for the Production of an Immunogenic Composition  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:

ADDRESS: Flinnegan, Henderson, Farabow, Garrett &  
ADDRESS: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington

```

? STATE: DC
? COUNTRY: USA
? ZIP: 20005-3315
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/468,057A  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/074,879  
FILING DATE: 16-JUN-1993  
APPLICATION NUMBER: WO 92/1136  
FILING DATE: 09-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E. R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 02356.0066-000000  
TELECOMMUNICATION INFORMATION:

? TELEPHONE: 202-408-4000  
 ? TELEFAX: 202-408-4400  
 ? INFORMATION FOR SEQ ID NO: 9:  
 ? SEQUENCE CHARACTERISTICS:  
 ? LENGTH: 7833 base pairs  
 ? TYPE: nucleic acid  
 ? STRANDEDNESS: single  
 ? TOPOLOGY: linear  
 ? MOLECULE TYPE: DNA (genomic)

Query Match	1.38; Score 35.2; DB 1; Length 7833;
-------------	--------------------------------------

Best Local Similarity 58.7%; Pred. NO. 1.8;  
Matches 61; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

```

QY      333  TACACTGATAAAAATCTGTCGCTACACAGGATTTCCAGCTTACTTTCTTAGCGGGCCC 392
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      5780  TCCTACTGTTAATATCTACAGCTGCCAGCAATAATATATATGCGCTGCGGTAAACATA 5722

```

```

QY      393 ATCATCGGTAATCACAAACCCCTCAGGAAAAGTGCACTTAAT 436
          ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      5720 ATCATCAGTATTGACAACTTCGCCACAGAAAGTGGAAGCAAAAT 5677

```

RESULT 7  
US-08-472-172-5  
; Sequence 5, Application US/08472172  
; Patent No. 5985288

```

1 GENERAL INFORMATION:
2 APPLICANT: Munson, Jr., Robert S
3 APPLICANT: Grass, Susan
4 APPLICANT: Chong, pele y y
5 APPLICANT: Fahim, Raafat
6 APPLICANT: Sia, Charles D y
7 APPLICANT: McVerry, Patrick
8 APPLICANT: Klein, Michel
9 TITLE OF INVENTION: Outer Membrane Protein P1 and Peptidase
10 TITLE OF INVENTION: Of Haemophilus Influenzae Type B
11 NUMBER OF SEQUENCES: 26
12 CORRESPONDENCE ADDRESS:
13

```

```

: ADDRESS: Slim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
:
:
:

```

```

? ZIP: MSG. IRO
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08,472,172
? FILING DATE: 07-JUN-1995

```

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/849,411  
FILING DATE: 07-JUL-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-4-71

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
TELEX: 063-24567 SIMBAS  
INFORMATION FOR SEQ. ID NO.: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1347 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

```

;          TOPOLOGY:      linear
US-08-472-172-5

Query Match           1.2%   Score 3
Best Local Similarity 49.2%   Pred. N
Matches              90;    Conservative 0; Mismatch 0
```

Query Match	1.28;	Score 34.2;	DB 2;	Length 1347;
Best Local Similarity	49.28;	Pred. No. 1.2;		
Matches 90;	Conservative 0;	Mismatches 93;	Indels 0;	Gaps 0;

OY 1018 GACATTACCTTCAATGGGAATGCCATTGTTCGACACTACCCAAACTGCAAAAAGAAAT 10777  
|| || | ||| ||||| ||| ||| ||| ||| ||| ||  
Db 223 GATTCTTGAAATTAAATGGAATGGTAGTGTGTAAGCTGCTTATATAAAAAGCTCCTCAATGAAC 282

1078 TCTATTGACATAGGATCTACTGCAAGATCAGCAATTACGTGCAATATCTGGGCATAGC 1137

```

Db 283 ACAAGAAAGACGGCTGACGCTTACAGAGCGTATGTTCTCGTGGCTTTTGTCGAAT 342
Oy 1138 AFTCTTTCTGACGATCGATTCGCTAATACGGCTCGGATTCACAGATCTTAAAT 1197
Db 343 CTTATTTCTGCTGCTCCAGTGAATGATTAATAGCGGTGCGAGGAATGATTAAT 402
Oy 1198 CTC 1200
Db 403 TTC 405

RESULT 8
US-07-876-280-5
; Sequence 5, Application US/07876280
; Patent No. 5262158
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M.
; APPLICANT: Cannon, Raymond J.C.
; APPLICANT: Bagley, Angela L.
; TITLE OF INVENTION: No. 5262158el Bacillus thuringiensis Isolates for
; NUMBER OF INVENTION: Controlling Acarides
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/876,280
; FILING DATE: 19920430
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: M/S 104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5600
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3771 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: 33f2
; IMMEDIATE SOURCE:
; CLONE: E. coli NM522(PMYC 2316) B-18785
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 4..24
; OTHER INFORMATION: /function- "oligonucleotide
; OTHER INFORMATION: hybridization probe"
; OTHER INFORMATION: /product- "GCA/T ACA/T TTA AAT GAA GTA/T TAT"
; OTHER INFORMATION: /standard_name- "probe a"
; OTHER INFORMATION: /note- "Probe A"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 13..33
; OTHER INFORMATION: /function- "oligonucleotide
; OTHER INFORMATION: hybridization probe"

```

```

; OTHER INFORMATION: /product- "AAT GAA GTA/T TAT CCA/T GTA/T AAT"
; OTHER INFORMATION: /standard_name- "Probe B"
; OTHER INFORMATION: /label- probe-b
; OTHER INFORMATION: /note- "probe b"
US-07-876-280-5

Query Match 1.2% Score 34.2; DB 1; Length 3771;
Best Local Similarity 50.3%; Pred. No. 2.4;
Matches 84; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Oy 1409 TTATGATGCGGCACACCTTAAGACATACAGAGGCTACCTTAACAGGCTTT 1468
Db 155 TTTTATGATGCTGACGACTTATATTAATACAGAGGAATGACCTTTAATATTA 214
Oy 1469 CCATTCCTGTAAGCTTTTACGAGGTAAGAAAGTTGTAATGCTGCTTCGAGCAA 1528
Db 215 AAGGTGTTTATCTGTATTAATCTTAATTCCTGAAGTTGTAATGCTGCTGAGCAA 274
Oy 1529 GTAAATATGATGACCTTATGCTGCGATTCCTTTTGATACCAA 1575
Db 275 GTACTATGTAAGTTTATTTGGCTTAAATATTTGGAGATTAACCA 321

RESULT 9
US-08-049-783-1
; Sequence 1, Application US/08049783
; Patent No. 5439881
; GENERAL INFORMATION:
; APPLICANT: Narva, Kenneth E
; APPLICANT: Schwab, George E
; APPLICANT: Payne, Jewel M
; TITLE OF INVENTION: Gene Encoding No. 5439881el Nematode-Active
; TITLE OF INVENTION: Toxins Cloned from Bacillus thuringiensis Isolates
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,783
; FILING DATE: 19930419
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1 (PS33f2):
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3771 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: 33f2
; IMMEDIATE SOURCE:
; CLONE: 33f2a
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 4..24

```



	Query Match	1.2%	Score 34.2;	DB 1;	Length 3771;
	Best Local Similarity	50.3%;	Pred. No. 2.4;		
	Matches 84;	Conservative 0;	Mismatches 83;	Indels 0;	Gaps 0
QY	1409	TTATGATGCGGGACACAGTTAAAGCAGACAGAGAGGCTCTTAAACAGGCTTT	1468		
Db	155	TTTTCAGATGTTTCATGCGACTTATATTAACACAGGAGAAATTGACCTTAAATGTATTA	214		

```

OY      1469 CCATTCCTGGAGACTCCTTTTAGCGGACGAGTAAGAATAATTGTAAATTCCTGCAGCAA 1528A
Db      215 AAGTGTTTTTATCTGATTTTAATCTTAATTCCTGAATTTGTACTGTGCGCTCTGCAGCAA 274
OY      1529 GTAAAAATGTAAGCCCTAGTGTCGCAATTCCTTTGGATAACAA 1575
Db      275 GTACATATTGTAAGTTTATTTATTGTCGCTAAAATATTGTGAGATAAACCA 321

RESULT 12
US-08-316-901A-5
; Sequence 5, Application US/08316301A
; Patent No. 5753492
GENERAL INFORMATION:
APPLICANT: Schepf, Harry E.
APPLICANT: Schwab, George E.
APPLICANT: Payne, Jewel M.
APPLICANT: Narva, Kenneth E.
APPLICANT: Fonceerrada, Luis
TITLE OF INVENTION: No. 5753492el Nematode-Active Toxins and Genes
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316, 301A
FILING DATE: 30-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/871, 510
FILING DATE: 23-APR-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/693, 018
FILING DATE: 03-MAY-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/565, 544
FILING DATE: 10-AUG-1990
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/084, 653
FILING DATE: 12-AUG-1987
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/830, 050
FILING DATE: 31-JAN-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lioyd, Jeff
REGISTRATION NUMBER: 35,589
REFERENCE/DOCKET NUMBER: MA20CCD1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 5 (PS3JF2):
SEQUENCE CHARACTERISTICS:
LENGTH: 3771 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

```





```

RESULT 14
US-09-173-891-5
: Sequence 5, Application US/09173891
: Patent No. 6077937
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel
: APPLICANT: Kennedy, M. Keith
: APPLICANT: Randall, John Brooks
: APPLICANT: Meler, Henry
: APPLICANT: Vick, Heidi Jane
: APPLICANT: Fonceriada, Luis
: APPLICANT: Schneckf, H. Ernest
: APPLICANT: Schwab, George E.
: APPLICANT: Fu, Jenny
: TITLE OF INVENTION: No. 6077937el Bacillus thuringiensis Toxins Active
: TITLE OF INVENTION: Against Hymenopteran Pests
: NUMBER OF SEQUENCES: 51
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David R. Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/173,891
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/158,232
: FILING DATE:
: APPLICATION NUMBER: US 07/887,980
: FILING DATE: 22-MAY-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/797,645
: FILING DATE: 25-NOV-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/703,977
: FILING DATE: 22-MAY-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: M/SC104.C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3771 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Bacillus thuringiensis
: INDIVIDUAL ISOLATE: 33P2
: IMMEDIATE SOURCE:
: CLONE: E. coli NM522(pMTC2316) B-18785
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 4..24
: OTHER INFORMATION: /function= "oligonucleotide
: OTHER INFORMATION: hybridization probe"
: OTHER INFORMATION: /product= "GCA/T ACA/T TTA AAT GAA GTA/T TAT"

```

```

OTHER INFORMATION: /standard_name= "probe a"
OTHER INFORMATION: /note= "Probe A"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 13..33
OTHER INFORMATION: /function= "oligonucleotide hybridization probe"
OTHER INFORMATION: /product= "AAT GAA GTA/T TAT CCA/T GTA/T AAT"
OTHER INFORMATION: /standard_name= "Probe B"
OTHER INFORMATION: /label= probe-b
OTHER INFORMATION: /note= "probe b"
US-09-173-891-5

Query Match      1.28; Score 34.2; DB 3; Length 3771;
Best Local Similarity 50.3%; Pred. No. 2.4;
Matches 84; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

OY 1409 TTATGATGCCGGCAGCACACCTTAAGAAGTATACAGAGAGGCACATTATAAGGTCTTT 1468
Db 155 TTTTAGATGTCGATCGATGGCATTAATTAACAAGAGGAATTGACCCTTAATGTAATTA 214
OY 1469 CCATTCCTGTCGACTCTTTAGGCGAGGGTAAAGATGTGTAATTCGCTTCGACAGCAA 1528
Db 215 AAGGCTTTTATCGTANTTAACCTTAATTCCTGAAAGTTGGTACTGTGGCCCTCGACAGCAA 274
OY 1529 GTAAAAATGTAGCCCTTAGTGTGTCGATTCCTCTTTTGATTAACCAA 1575
Db 275 GTACTATGTGAAGTTTATTTGCGCTAAAATATTTGGAGATTAACCA 321

RESULT 15
US-09-076-137-5
Sequence 5, Application US/09076137B
Patent No. 6166195
GENERAL INFORMATION:
APPLICANT: Schepf, Harry E.
APPLICANT: Schwab, George E.
APPLICANT: Payne, Jewel M.
APPLICANT: Narva, Kenneth E.
APPLICANT: Foncerrada, Luis
TITLE OF INVENTION: No. 6166195el Nematode Active Toxins and Genes Which Code
TITLE OF INVENTION: Therefore
FILE REFERENCE: NA-20CCCD2
CURRENT APPLICATION NUMBER: US/09/076.137B
CURRENT FILING DATE: 1998-05-12
EARLIER APPLICATION NUMBER: 08/316,301
EARLIER FILING DATE: 1994-09-30
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 3771
TYPE: DNA
ORGANISM: Bacillus thuringiensis
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4)..(24)
OTHER INFORMATION: /function= "oligonucleotide hybridization probe"
OTHER INFORMATION: /product= "GCA/T ACA/T TTA AAT GAA GTA/T TAT"
OTHER INFORMATION: /standard_name= "probe a" /note= "Probe A"
FEATURE:
NAME/KEY: misc_feature
LOCATION: (13)..(33)
OTHER INFORMATION: /function= "oligonucleotide hybridization probe"
OTHER INFORMATION: /product= "AAT GAA GTA/T TAT CCA/T GTA/T AAT"
OTHER INFORMATION: /standard_name= "Probe B" /label= probe-b /note=
US-09-076-137-5

Query Match      1.28; Score 34.2; DB 4; Length 3771;
Best Local Similarity 50.3%; Pred. No. 2.4;
Matches 84; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```

Tue Oct 9 10:45:11 2001

us-09-446-677b-3.rni

Page 10

QY 1409 TTATGATGCGGGGACACACTTAAAGCAAGTACGAGAGAGTGCTACTTTACAGTCTTT 1468  
 ||| ||||| ||| ||| ||||| ||| |||  
 Db 155 tttagatgtgtcagtgactatataactacagagaaattaccctttaaagtlaata 214  
 ||| ||||| ||| ||| ||||| ||| |||  
 QY 1469 CCATTCCTGTTAGACTCTTTTGCGGAGAGGTATAGAAAGTTGTTATTCCTGCTTCGAGCAA 1528  
 ||| ||| ||| ||||| ||| ||| ||||| ||| |||  
 Db 215 aaggtgttatacttgtaacttaacttaattcccgaaagtgtgtacgcgtgyccttcgaacaa 274  
 ||| ||| ||| ||| ||||| ||| ||| ||||| ||| |||  
 QY 1529 GTAAAACTTAGCCCTTATGAGGTCCCATTCCTCTTTTGGATACCAA 1575  
 ||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| |||  
 Db 275 gtaactatgtgaagtttaatttgcgcctcaaatatatttggagataaccaa 321

Search completed: October 2, 2001, 03:26:18  
Job time: 50171 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2001, 03:06:06 ; Search time 10300.6 Seconds  
(without alignments)  
2583.331 Million cell updates/sec

Title: US-09-446-677B-3  
Perfect score: 2815  
Sequence: 1 ATGAATGCGCAATTTCCTG.....CTCATGTCACAGAAATTCG 2815

Scoring table: IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 1022815 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:\*

1:	gb_est1:**	44:	em_esthum10:**
2:	gb_est2:**	45:	em_esthum11:**
3:	gb_est3:**	46:	em_esthum12:**
4:	gb_est4:**	47:	em_esthum13:**
5:	gb_est5:**	48:	em_esthum14:**
6:	gb_est6:**	49:	em_esthum15:**
7:	gb_est7:**	50:	em_esthum16:**
8:	gb_est8:**	51:	em_esthum17:**
9:	gb_est9:**	52:	em_esthum18:**
10:	gb_est10:**	53:	em_esthum19:**
11:	gb_est11:**	54:	em_esthum20:**
12:	gb_est12:**	55:	em_esthum21:**
13:	gb_est13:**	56:	em_esthum22:**
14:	gb_est14:**	57:	em_esthum23:**
15:	gb_est15:**	58:	em_esthum24:**
16:	gb_est16:**	59:	em_esthum25:**
17:	gb_est17:**	60:	em_esthum26:**
18:	gb_est18:**	61:	em_esthum27:**
19:	gb_est19:**	62:	em_esthum28:**
20:	gb_est20:**	63:	em_estin1:**
21:	gb_est21:**	64:	em_estin2:**
22:	gb_est22:**	65:	em_estin3:**
23:	gb_est23:**	66:	em_estin4:**
24:	gb_est24:**	67:	em_estin5:**
25:	gb_est33:**	68:	em_estom1:**
26:	gb_est34:**	69:	em_estom2:**
27:	gb_est35:**	70:	em_estov1:**
28:	gb_est36:**	71:	em_estov2:**
29:	gb_est37:**	72:	em_estp11:**
30:	gb_est38:**	73:	em_estp12:**
31:	gb_est39:**	74:	em_estp13:**
32:	gb_est40:**	75:	em_estp14:**
33:	em_estba:**	76:	em_estp15:**
34:	em_estfun:**	77:	em_estp16:**
35:	em_esthum1:**	78:	em_estp17:**
36:	em_esthum2:**	79:	em_estp18:**
37:	em_esthum3:**	80:	em_estp19:**
38:	em_esthum4:**	81:	em_estp110:**
39:	em_esthum5:**	82:	em_estro1:**
40:	em_esthum6:**	83:	em_estro2:**
41:	em_esthum7:**	84:	em_estro3:**
42:	em_esthum8:**	85:	em_estro4:**
43:	em_esthum9:**	86:	em_estro5:**
		87:	em_estro6:**
		88:	em_estro7:**
		89:	em_estro8:**
		90:	em_estro9:**
		91:	em_estro10:**
		92:	em_estro11:**
		93:	em_estro12:**
		94:	em_estro13:**
		95:	em_estro14:**
		96:	em_estro15:**
		97:	em_estro16:**
		98:	em_estro17:**
		99:	em_estro18:**
		100:	em_estro19:**
		101:	em_estro20:**
		102:	gb_est25:**
		103:	gb_est26:**
		104:	gb_est27:**
		105:	gb_est28:**
		106:	gb_est29:**
		107:	gb_est30:**
		108:	gb_est31:**
		109:	gb_est32:**
		110:	gb_est41:**
		111:	gb_est42:**
		112:	gb_est43:**
		113:	gb_est44:**
		114:	gb_est45:**
		115:	gb_est46:**
		116:	gb_est47:**

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



DB	707	TACTGCTACTGCTACTACTACTGCTACTGCTACTGCTACTACTACTACTACTGCT	766
OY	1212	TGC 1214	
DB	767	TAC 769	
RESULT	2		
LOCUS	AN061466		
DEFINITION	AN061466 579 bp mRNA EST 20-MAY-1999		
ACCESSION	AN061466		
VERSION	AN061466.1 GI:4882570		
KEYWORDS	EST.		
SOURCE	Dictyostelium discoideum.		
ORGANISM	Dictyostelium discoideum.		
REFERENCE	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.		
AUTHORS	1 (bases 1 to 579)		
TITLE	Yoshino, R., Morio, T. and Tanaka, Y.		
JOURNAL	Developmental CDNA in Dictyostelium discoideum		
COMMENT	Unpublished (1997)		
	Contact: Hideko Urushihara		
	Institute of Biological Sciences		
	University of Tsukuba		
	3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan		
	Email: d402hu@tsakura.cc.tsukuba.ac.jp		
	PROJECT = Dictyostelium discoideum CDNA project in Japan.		
	Location/Qualifiers		
	1..579		
	/organism="Dictyostelium discoideum"		
	/strain="AX4"		
	/db_xref="taxon:44689"		
	/clone="SLE356"		
	/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"		
	/dev_stage="slug"		
	/dev_stage="slug"		
BASE COUNT	138 a 138 a 225 g 136 t 1 others		
ORIGIN			
Query Match	1.6%; Score 45.6; DB 107; Length 579;		
Best Local Similarity	45.4%; Pred. No. 0.074;		
Matches	248; Conservative 0; Mismatches 289; Indels 9; Gaps 2		
OY	430	GTTAATGTCGAGGGGAGCTTACATTGATTAACAATGCAGACTTATTTATTTAAACAAGT	489
DB	41	GCTAACGCTGAGGGGCTACTGAGAGAAATACGGCGGTCTTCATCACTGCTGGCAGT	100
OY	490	TACTGTGAGGAAATAGGGGAGGACATTTCTACCAAGAAATCTTTTGAAGAAACAGCAG	549
DB	101	ACTGAGAGTAATAGGGGGGCTCTACTGAGAGTAATAGGGGGGCTCTACTGAGAGCAAT	160
OY	550	GGATCGATTCTTTTGAAAGGCAATTAATCGAGCGCAACAGGAGAAAGTGGGGCTATT	609
DB	161	AGTGAAGCTGCTACAGGGGTGCTACAGTGGAGGCTGCTACTGAGGAGAAATGAGAGGTAAC	220
OY	610	TGTGCTACTGCTGATGATTAATTAACAATTAATAGGGCTCCACACCTCTTTCGAACAT	669
DB	221	AGTGTGCTCTATCTTTCAGAGAGTAATAGCGGTGGTTTAAACAGGTGCTCTACTGTGCT	280
OY	670	ATTGCTGAGAGCTGACAGGTGAGAGCTTAATAATACACAGGAACTGTACAAATTACAGGGAAT	729
DB	281	AATAGTGAAGGTGCTACTGAGAGCAATATGTTGGCGCTGCACACCGGAGGAAATATGTTGCT	340
OY	730	ACGTCTCTTTGATTTTTCGAAATATAGTGTGACAGGAGCGCAGGAGAAATGGAGAGCTTT	789
DB	341	ACTGTGCTTAATAGAGGAGGAAATAGTGGCGCTGCACACCGGAGGAAATATGAGAGTGTCT	400
OY	790	TCTGAGATGCGGATGTTACCATATCTGAGGATTCAGAGTGAATCTTTCAGGAAACCA	849
DB	401	TCAAGTGGTTCAGTCTT-----TCAGAGAGTAATAGCGGTGGTTTAAACAGGAGCGCT	454
OY	850	GCTGTAGCTAATGGGAGCAATTTATGCTAAGAAAGCTTACAGTGGCTTCGGGGGGGGG	909

DB	455	ACTGAGGTAATAGTGGCGGTCTACTGAGGTAATAGTGGCGGTCTACTGAGGTAAT	514
OY	910	GGGGTATCTCTTTTCTTAACAATATAGTCCAGGTAACCACTGCAGAGTATGAGGCC	969
DB	515	AGTGGCGGTCTACTGAGGTAATAGTGGCGGTCTACTGAGGTAATAGTGGCGGTCT	571
OY	970	ATTTCG 975	
DB	572	GCTACT 577	
RESULT	3		
LOCUS	AU060774		
DEFINITION	AU060774 Dictyostelium discoideum SL (H. Urushihara) Dictyostelium		
ACCESSION	AU060774		
VERSION	AU060774.1		
KEYWORDS	EST.		
SOURCE	Dictyostelium discoideum.		
ORGANISM	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.		
REFERENCE	1 (bases 1 to 623)		
AUTHORS	Yoshino, R., Morio, T. and Tanaka, Y.		
TITLE	Developmental cDNA in Dictyostelium discoideum		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Hideko Urushihara Institute of Biological Sciences University of Tsukuba 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan Email: d402huesakura.cc.tsukuba.ac.jp PROJECT = Dictyostelium discoideum cDNA project in Japan. Location/Qualifiers 1. .623 /organism="Dictyostelium discoideum" /strain="AX4" /db_xref="taxon:44689" /clone="SLB441" /clone_lib="Dictyostelium discoideum SL (H. Urushihara)" /dev_stage="slug"		
FEATURES			
source			
BASE COUNT	168 a 104 c 214 g 136 t 1 others		
ORIGIN			
Query Match	1.6%	Score 45.2; DB 107; Length 623;	
Best Local Similarity	46.5%; Pred. No. 0.099;		
Matches 146; Conservative 0; Mismatches 168; Indels 0; Gaps 0;			
OY	550	GGATGATTTCTTTTGAAGGAATAATCGACCGCAGAGGAAAAAGTGGGCTATT	609
DB	107	GTTTCAGTCTCTTCAGGAGGTATATGCGGTTGTTAACAAGAGGCGCTACTGAGATAT	166
OY	610	TGTGCTACTGCTAGTATATTAACAATAATACGGCTCCTACCTCTTCGAACAT	669
DB	167	AGTGGCGGTCTACTGAGGTAATAGTGGCGGTCTACTGAGGTAATAGTGGCGGTCT	226
OY	670	ATTGCTGAACCTGCAGGTGAGCTTAATATGCAACAGAACTGTACATTTACAGGAT	729
DB	227	ACTGAGGTAAATAGTGGAGGTCTACTGAGGTAATAGTGGAGGCTACTGCTGTAT	286
OY	730	ACGTCCTTTGTAATTTTCGAAAAATAGTGTGACAGCGACCGCAGAAATGAGAGACTCT	789
DB	287	AGTGTGCTGCACACGAGCAAAATAGTGTGTGTCGAACGAGGCAAAATAGTGGCGTGA	346
OY	790	TCTGAGATGCCCGATGTTACCATATCTGGGAATCAAGAGTGAATCTTCTCAGAAAA	849
DB	347	ACTGAGCAAAATAGTGGGTCTGCAACTGTGCAATATGGAACAATGCCCTGTAAT	406
OY	850	GCTGACTAATGG 863	
DB	407	GGTGCTGCAATGG 420	





[illegible]

AUTHORS	Genoscope.
TITLE	Direct Submision
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/tetraodon">http://www.genoscope.cns.fr/tetraodon</a> .
FEATURES	location/Qualifiers
SOURCE	1..989 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /cclone="139021" /clong="lib-g" /note="Genoscope sequence ID : COAG139AH11np1-end : T7"
BASE COUNT	320 a     245 c     122 g     261 t     41 others
ORIGIN	
Query Match	1.5%; Score 42.4; DB 220; Length 989;
Best Local Similarity	40.6%; Pred. No. 0.76;
Matches	203; Conservative 12; Mismatches 285; Indels 0; Gaps 0;
Oy	1019 ACATTACCTTCATGGGAATGCGATTGTTCGCACTCACACCACAACCTCAAAAGAATT 1078
Dd	82 ATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 141
Oy	1079 CTATTGACATGAGATCTACTCCAAAAGATCAGCAATTTACGTCAATATCTGGCATGCA 1138
Dd	142 CTACGACTACTAAATCTACTACTATACTACTAAATACGACTAAATCTACTACTACTA 201
Oy	1139 TCTTTTTCAGATCCGATTACTGCTTAATACGCCTGCCGATTTACAGATCTTTAATC 1198
Dd	202 ATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 261
Oy	1199 TCATTAAGGCTGATGACAGTAAATAGATTAAGTGAGGGTCAGTTGTTTTCTGGTG 1258
Dd	262 CTCAGCTGCTTAATCTACTACTATACTACTACTACTACTACTACTACTACTACTACTA 321
Oy	1239 AAAAGCTCTCTGAGATGAGCAAAAAAGTTGCAGACACCTCACTTCTAGCTGAAGCAGC 1318
Dd	322 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 381
Oy	1319 CTGTAACTCTAACGACAGAAATTTAGTAACTTAAACCGGTGTCACCTGCATCGAAAG 1378
Dd	382 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTATA 441
Oy	1379 GCTTACTACAGACCGGGGCTCTCTGTTATTATATGATGCGGGACACAAGTTAAAGCAA 1438
Dd	442 ATAATAAATAAATNNNTNGTGTGGKGGTGGGKGKTCDTACTACTACTACTACTATA 501
Oy	1439 GTACGAGAGAGTCACTTTAACAGGTCTTTCCATTTCTGTAGACTCTTTAGCGAGGTA 1498
Dd	502 AATAATAAATAAATAATMTACTACTCCCTCCATCCCCCCCCCGCYGCRGGYGGTG 561
Oy	1499 AGAAGTTGTATATGCTGCT 1518
Dd	562 TGTCGGGGGGGTGTGTGTGT 581
RESULT 8	
CNS0106X	GSS
LOCUS	CNS0106X 1101 bp DNA 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL098595
VERSION	AL098595.1 GI:5610206
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Plasmod Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

JOURNAL		Published (1997)	
COMMENT	CONTACT	Institute of Biological Sciences University of Tsukuba 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan Email: d402huesakura.cc.tsukuba.ac.jp PROJECT = Dictyostelium discoidium cDNA project in Japan.	
FEATURES	SOURCE	Location/Qualifiers 1..558 /organism="Dictyostelium discoidium" /strain="AX4" /db_xref="taxon:44689" /clone="SIG686" /clone.lib="Dictyostelium discoidium SL (H.Urushihara)" /dev_stage="slug"	
BASE COUNT	133 a 76 c 215 g 134 t	ORIGIN	
Query Match	1.5%; Score 41.8; DB 107; Length 558; Best Local Similarity 45.1%; Pred. No. 0.93; Mismatches 287; Indels 9; Gaps 2;		
Matches: 243; Conservative	0; Mismatches		
QY	437 GTGAGGGGATCTTACATTTGATTAACAATGGAACATATTATTTAAACAGATTACTGTG	496	
Db	3 GTGGAGGCTACTCTGAGGAATACCGCGGCTTTCATCATCTGTGGCAATACTGGAG	62	
QY	497 AGGAAATGGCGGAGCCATTTCTACCAAGATCTTCTTTAAAAACAGCAGGATCGA	556	
Db	63 GTAATAGCGCGGTCTACTGGAGTAATACCGCGGTCTACTGGAGCAATAGTGAG	122	
QY	557 TTTCTTTTGAAGGGAATTAATCGACCGCAGGAAAGAAAGTGGCGCTATTGTGCTA	616	
Db	123 GTGCTACAGGTGTACAGCTGGAGTGTCTACTGAGGAATATGTGAGTACAGTGTG	182	
QY	617 CTGCTACTGTAGATTACAAATAATACGCTCCTACCTCTTCTGCAACAATATTGCTG	676	
Db	183 CCTCATCTTCAGAGGTATAGCGTGTGTTAACAGGTGGTGTCTACTGTGTATATAGT	242	
QY	677 AAGCTGACGAGGAGCTTAATAATGACAGGAACCTGATACATTAACGGAATACGTCTC	736	
Db	243 GAGGTCTACTGAGGAATATGTGCGCTGCAACCGGAGGAATATGTGTCTACTGTGTG	302	
QY	737 TTGTATTTTCGAAATATGTGTGACAGCGACCGCAGGAATATGAGAGCTCTTCTGGAG	796	
Db	303 GTAATAGTGGAGGAATATGTGCGCTGCAACCGGAGGAATATGTGAGTGTCTCAAGTG	362	
QY	797 ATGCGGATTTACCATATCTGGGAATCAGAGTGTACTTTCTCAGAGAAACCAAGCTGTAG	856	
Db	363 GTTCCAGTTCTTCAAGGTATTAATACCGTGTGTTAATGAGGAGCCCTACTGTGAG	416	
QY	857 CTAATGGCGGAGCCATTATGCTAAGAAGCTTACACTGGCTCCGGGGGGGGGGGGGCTA	916	
Db	417 GTAAATAGTGGGTGTCTACTGGAGTAAATATGTGGCGGTCTACTGTGAGGTAAATAGTGCGG	476	
QY	917 TCTCCTTTTCTAACAAATATAGTCCAGGTACCACTGCAAGTAAATGTGTGAGCCATTCT	975	
Db	477 GTGTACTGAGAGTAAATAGTG---GAGGTGTCTACTGAGGTAAATAGTGAGGTCTACT	532	
RESULT 10	FR0006944		
LOCUS	FR0006944 619 bp DNA		
DEFINITION	F. rubripes GSS sequence, clone 13B16aCT, genomic survey sequence.		
ACCESSION	290754		
VERSION	290754.1 GI:1867968		
KEYWORDS	GSS: genome survey sequence.		
SOURCE	Takifugu rubripes.		
ORGANISM	Takifugu rubripes.		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes; Tetraodontidae; Takifugu		



```

/organism="Entamoeba histolytica"
/strain="HML:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOSt1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G. and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
light size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaundin and B.
Barell, Oxford University Press, 1999)."
```

Query Match	1.5%: Score 41; DB 245; Length 887;
Best Local Similarity	46.9%: Pred. No. 1.9; Mismatches 128; Conservative 0; Indels 0; Gaps 0;
Matches 128; Conservative 0; Mismatches 145; Indels 0; Gaps 0;	
Oy 1022 TTACCTTCATGGGAAATGCCATTGTGGCACTACACACAACAATACAAAAGAAATTCCTA	1081
Db 566 TTACATTTAGAGGGGCAATTACATATTTTATACACACAAATCCACCATCATTAATTC	507
Oy 1082 TTGCATATGAGATCTACTGTCCAAAGATCAGCATTTTACGTGCATATCTGGGATATGACATCT	1141
Db 506 TTCAATATTCACATCACAATAATTTTAAAAATTACACATCCATCGCTGTGATAGTCTT	447
Oy 1142 TTTTCTACATCCGATTCATCTGCTTAATACGGCTGGGATTTTACAGATACCTTAATTCCTCA	1201
Db 446 ATTTCGGCTTATCCACTTATGACATTTCCCTCATCATATAGTTGAACCTCCACAAAATGCTG	387
Oy 1202 ATAAAGCGTATGACAGGTATATAGTACAGATTTTAAAGGGCTGCATTTTCTTCTGGTGA	1261
Db 386 CAGCATATACACAGAAACAATTAACAATTTCAATTAATGTGATTTCATGAGAGTAATAAA	327
Oy 1262 AGCTCTCTGAAGATGAGCAAAAAGTTCGACACA	1294
Db 326 ATATTATGAGACCAACTCAACCCACAAGAA	294

RESULT	13			
AM465899/c				
LOCUS	AM465899	520 bp	mRNA	EST
DEFINITION	BP230020B20H7 Soares normalized bovine placenta		Bos taurus	CDNA
ACCESSION	clone BP230020B20H7 5',			mRNA sequence.
VERSION	AM465899			
KEYWORDS	AM465899.1	GI:7036067		EST.
SOURCE				
ORGANISM	cow.			
REFERENCE	Bos taurus			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.			
TITLE	1 (bases 1 to 520)			
JOURNAL	Lewin, H.A., Soares, M.B., Rebelz, M., Pardinas, J., Liu, L. and Larson, J.H.			
COMMENT	Bovine ESTs			
	Unpublished (2000)			
	Contact: Lewin, H. A.			
	W. M. Keck Center for Comparative and Functional Genomics			
	University of Illinois at Urbana-Champaign			
	340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL			
	61801, USA			
	Tel: 217 333 5998			
	Fax: 217 244 5617			
	Email: h-lewin@uiuc.edu			
	Funding for cattle EST sequencing was provided by the USDA National			

```

RESEARCH INITIATIVE, ANIMAL GENOME RESOURCE GRANT AG 99-3205-8534
TO H. A. LEWIN AND J. E. MOMACK. BASE CALLING/QUALITY SCORES: PHRED
FROM WASHINGTON UNIVERSITY GENOME CENTER. VECTOR TRIUMAL 9:
CROSS_MATCH FROM WASHINGTON UNIVERSITY GENOME CENTER PHRAP SUITE.
SEQUENCES SUBMITTED ARE VECTOR FREE AND AT LEAST 200 BP IN LENGTH.
PCR PRIMERS
FORWARD: TAATAGACATCAGTATAGGG
BACKWARD: ATTAACCCCTCAGTAAG
Insert Length: 520 Std Error: 0.00
Plate: BP230020B20 Row: H Column: 7
Seq primer: AGCGATATACAACTTTCACACGAGA
High quality sequence stop: 520.
location/Qualifiers
1..520
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="BP230020B20H7"
/clone_1lb="Soares normalized bovine placenta"
/sex="female"
/lab_host="DH10B"
/note="Organ: Placenta; Vector: pT73Pac; Site: 1: EcoRI;
Site_2: NotI; The cDNA library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806. "
191 a 85 c 160 g 84 t

```

	Query Match	Similarity	1.4%;	Score 40.8;	DB 116;	Length 520;
	Best Local	Similarity	53.0%;	Pred. No. 1.8;		
	Matches	87;	Conservative	0;	Mismatches 77;	Indels 0;
						Gaps 0;
Qy	1630	CAGCTCTGCTGCTGGTACTGCAACACACTACACATGTGTTCCAGCGGTTCTTACATAGCA	1689			
Db	240	CCGGCCCTACTCCGACTTCTGCTCTACTACTACTGTAAGTGGTGGCTGCTACTCTCTCT	181			
Qy	1690	ACTCCTACGCACTATGGTATTCACAGTACTTGGGGAATGACTTGGGTTGATGATACCGCA	1749			
Db	180	ACTGCTGTGCACATTTGACTTGAAGTCTCTACTGSAAGTACTGCACTACATGCTACTGCC	121			
Qy	1750	AGCACTCCAAAGACTTAAGACAGCGACGACANTTACTTGGACCAATAC	1793			
Db	120	ACTGCTGGAACTGTCTTCCACTGCTGCTACTGGTTCAACTGCTAC	77			

RESULT	14
FR0043207	
LOCUS	
DEFINITION	FR0043207 550 bp DNA GSS 22-OCT-1999 Fugu rubripes GSS sequence, clone 161G06ac7, genomic survey sequence.
ACCESSION	AL130699
VERSION	AL130699.1 GI:6112645
KEYWORDS	GSS; genome survey sequence. Takifugu rubripes.
SOURCE	Takifugu rubripes.
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu. 1 (bases 1 to 550) Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K., Umanita,Y., Williams,G. and Brenner,S.
REFERENCE	Direct Submission Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB. UK Email: biohelp@ngmp.mrc.ac.uk Vector: pBluescript II KS V.type: phagemid PRIMER: KS
AUTHORS	DESCR:
JOURNAL	One pass dye-terminator sequencing of cosmid cloned genomic sequence.
COMMENT	

FEATURES	Location/Qualifiers
source	1..550
	/organization="Takifugu rubripes"
	/db_xref="taxon:31033"
	/clone-lib="cosmid 161G06"
	/clone="161G06AC7"
BASE COUNT	145 a 155 c 82 g 163 t
ORIGIN	5 others

	Query Match	Similarity	1.4%	Score 40.8	DB 222	Length 550
	Best Local	Similarity	47.6%	Pred. No. 1.8	Mismatches 133	Indels 0
	Matches	120	Conservative	0	Gaps	0
QY	583	GCACAGGAAAAAGGTGGGGCTATTGTGCTACTGCTAGTCTGTAGATTTACAAATTAAT	642			
Db	112	GCTACTACTACTACTGTGTGCTGCTGCTACTGCTACTACTACTACTACTCCCTACTACT	171			
QY	643	ACGGCTCTACCCCTTCTTCGAAACATATTGCTGAAGCTCAGGTGGAGCTTTAAATAGC	702			
Db	172	GCTGTTACACCCGATTTGCTGCCACTGCTGCTGCTAGCTGCTGCTGCTACTACTACT	231			
QY	703	ACAGGAAATGTACATTTACAGGGAATACGTCTTTGTTTTCGAAATATGTTGACA	762			
Db	232	ACTACTACAGCTACTACTCAATTAATTTGCTGCTACTACTGCTGCTACTACTCTGCCACT	291			
QY	763	GCAGCCCGAGGAAATGAGAGAGCTTTTCTGAGATGCGATGTTACCATATCTGGGAAT	822			
Db	292	ACTGTTACAGCTAATATACGCTACTGCTCTATTGTTGCTGCTACTACTACTACTACT	351			
QY	823	CAGACTGTAACT	834			
Db	352	ACTACTACCCT	363			

RESULT	15
AU033711	
LOCUS	AU033711.607 bp mRNA EST "28-APR-1999
DEFINITION	AU033711 Dictyostelium discoidium SL (H.Urushihara) Dictyostelium
ACCESSION	discoidium cDNA clone SLB346, mRNA sequence.
VERSION	AU033711
KEYWORDS	AU033711.1 GI:3799135
SOURCE	EST.
ORGANISM	Dictyostelium discoidium.
REFERENCE	Dictyostelium discoidium
AUTHORS	Eukaryota; Mycetozoa: Dictyostellida; Dictyostelium. 1 (bases 1 to 607)
TITLE	Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mita,B.N., Pt,M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y. The Dictyostelium developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development
JOURNAL	DNA Res. 5 (6), 335-340 (1998)
MEDLINE	99156227
COMMENT	Contact: Hideko Urushihara Institute of Biological Sciences University of Tsukuba 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan Email: da02n@esakura.cc.tsukuba.ac.jp PROJECT = 'Dictyostelium discoidium cDNA project in Japan' POLYA-No.

[illegible]

Search completed: October 2, 2001, 03:06:19  
Job time: 51637 sec

```

FEATURES
source
1. 607
location/Qualifiers
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLB346"
/clone_lib="dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"
BASE COUNT
203 a 125 c 134 g 145 t
ORIGIN

```

BASE COUNT	203 a	125 c	134 g	145 t
ORIGIN				

BEST AVAILABLE COPY

**THIS PAGE BLANK (USPTO)**



```

RESULT 10
US-08-469-880-2
; Sequence 2, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstesser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MTS:V9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-880-2

Query Match 5.0%; Score 238; DB 2; Length 1536;
Best Local Similarity 21.18; Pred. No. 2.6e-09;
Matches 134; Conservative 99; Mismatches 235; Indels 168; Gaps 27;
QY 34 SFNGNID-----SGTFPKTSATYSLTGDF---FYEPKGTPLSDFSCFQOT----- 78
DB 736 SFNGNISVSGGSDVDFLLSSSSVNOIPGVYINSKYFNVSSTGSLR---FKTSSSTKTF 792
QY 79 --TNNLFLGHSGLTGFIDAGTHAGAASTANKULTFSGFGLSPDSSPSTVTGQ 136
DB 793 SIEDLTLNATAGNITLLOVE-ETDGMIGKGIYAKKNITFEKGN-ITFGSRKAVTEIEGN 850
QY 137 GTLSAGGVN-----ENIRKLVAAGNFSTADGAIKGASLLTGTSGDALFSNNSSTK 191
DB 851 VTINNNAVTLIGSDPDHOKPL-----TIKKDVIINSGLTA 888
QY 192 GGAATATAGARIANNTGTVRFSLNIASTSGAIDDEGTSILS-----NNKFLY 239
DB 889 GGNIVNIAGNLTVESNMFKAITNFTFNVGSLFPNKGNSNLSIAKGARFKIDNSKMLS 948
QY 240 FEGNAATTGGAICNTRKASGSPELLISNNKTLIFASNAETSGAIIAKK--IALSSGGF 297

```

```

DB 949 ITTSSSYPTTIGSINTNKNKGDNLITNE-----GSDTEMQIGDGVSOKEGNLTSSDKI 1003
QY 298 TEFIRNNVSSATPPGGA-----ISIDASELSLSEATGNTTVRN----- 337
DB 1004 ----NITKQITIRAGDGENSDDATNNANLTIKTRELKLTQDNLINSGFKAEITAKDG 1058
QY 338 ---TLTTGSDTDFKRAININSGNK---FTELR-----AAKNHILFDFDPTISGTSADV 387
DB 1059 SDLTIGTNSAD-----GTNAKVTFNQVAKDSKISALGHVYTLHSKVTSGSNNT 1109
QY 388 LKINGSAGALNPQGTILFSGETLTADELKVDNLKSSFTQPSVLSGKLLLOKGVTL 447
DB 1110 EDDSDNNAGL-----TIDAKNVTVNNNITSHKAVSISATSGEITFTKGTIN 1156
QY 448 STS-----FSQFAGSLG---MDSGT-PLSTAGSITTNLG-----INVDLGLKQPSL 494
DB 1157 ATTGNVEITAGTGSILGIESSSGSVTLTATEGALAVSNISGNTVTVYANGALTTLAGS 1216
QY 495 TAKGASNKVIVSGKLNLDIEGNIYESHMFSDQLFSLIKITVDADVDYTNDISLLPVP 554
DB 1217 TIKG-TESVTTSSQSG--DIGTISG-----TVEVKATESLITQSNKIK 1259
QY 555 AEDPNSEYGFQGMNVMNTDT-----ATNTEKAPA 585
DB 1260 AT-----TGEANVTSATGTIGTISGNTVNTYA 1287

RESULT 11
US-08-728-470-2
; Sequence 2, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstesser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids

```

MOLECULE TYPE: DNA (genomic)  
US-08-302-832-2

Query Match 5.0%; Score 238; DB 1; Length 1536;  
Best Local Similarity 21.1%; Pred. No. 2.6e-09;  
Matches 134; Conservative 99; Mismatches 235; Indels 168; Gaps 27;

```

OY 34 SFNGNID-----SGTTPKTSATYSLGDFV---FYEPGKGTPLSDSCFQOT----- 78
DB 736 SFNGNISVGGGSDVDFTLASSNVQTPGVINSKYFNVSTGSSLR--FKTSGTKTGF 792
OY 79 --TDNITFLGNHSLTFGIDAGTHAGAASATTANKNLTFSGFSLSPDSSSTVTTGQ 136
DB 793 STEKDLITNATGNTLLQVE-GTDMIGKGIYAKKNITFEGGN-ITFESRAVTEIEGN 850
OY 137 GTLSSAGVNL-----ENIRKLVAAGNFSTADGAIKGASFLTGTSGDALEFNNSSSTK 191
DB 851 VTINNANVTLLIGSDFDNHOKPL-----TIKKDVIINSGNLTA 888
OY 192 GGAITTAGARIANNNGYVRFSLNASTSGAIDDEGTSILS-----NNKFY 239
DB 889 GGNIVINAGNLTVESNANFKAITNFTFNVGGLFDKNGNSNISIAKGARFKDIDNSKNLS 948
OY 240 FEGNAKTGGAICMCTKASGSPDLIISNNKTLIFASNVAFETSGAIIHAKK--LALSOGF 297
DB 949 ITTNSSTYRTIISGNTINKNDLITNE-----GSDTEMQIGDVSQKEGNLTJSSDKI 1003
OY 298 TEFLNNVSSATPKGA-----ISIDASGELSLAETGNITFVRN----- 337
DB 1004 -----NITKOITIKAGVDGENDSDATNNANLITKTELKLPQDNLINISGFNKAETAKDG 1058
OY 338 ---TLTTGSDTTPKRNAINISNGK---FTELR-----AAKHHTFFPDPTISEGSSDV 387
DB 1059 SDLTIGNTNSAD-----GTNAKVTYFNQVKDSKISADHKYTLHKSKEVETSSNNNT 1109
OY 388 LKINNGSAGALNRYOGTILFSGETFLADELKYADNLKSFOTPOVSLSGKLLQKVTLE 447
DB 1110 EDDSDNNAGL-----TIDAKNVTYNNNITSHKAVISATISEITTKTGTTIN 1156
OY 448 STS-----FSQENGSLG---MDSGT-TLSTAGSITITNLG-----INVDISGLKOPVSL 494
DB 1157 ATTGAVEITAGQTSILGIESSSGVTLLATGALAVSNISGNFTVTVANSALTTLAGS 1216
OY 495 TAKGASNKVIVSGKLNLDIEGNIYESHMFSDQLFSLIKITVDADVDPNDISSLIPIV 554
DB 1217 TIKG-TESTVTSQSG--DIGGTISG-----TVEVKATESLTTQNSKTIK 1259
OY 555 AEDPNSYGFQOGAMVNMWTTDT-----ATNTKEATA 585
DB 1260 AT-----TGEANVTSATGTIGGTISGNTVNVTA 1287

```

RESULT 9  
US-08-530-198-2  
Sequence 2, Application US/08530198  
Patent No. 5869065

GENERAL INFORMATION:  
APPLICANT: BARENKAMP, STEPHEN J  
APPLICANT: ST. GENE III, JOSEPH W  
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESS: Shoemaker and Mattare, Ltd  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/530.198  
FILING DATE: 13-DEC-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: BERSKRESSER, JERRY W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: JWB-1186  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-530-198-2

Query Match 5.0%; Score 238; DB 2; Length 1536;  
Best Local Similarity 21.1%; Pred. No. 2.6e-09;  
Matches 134; Conservative 99; Mismatches 235; Indels 168; Gaps 27;

```

OY 34 SFNGNID-----SGTTPKTSATYSLGDFV---FYEPGKGTPLSDSCFQOT----- 78
DB 736 SFNGNISVGGGSDVDFTLASSNVQTPGVINSKYFNVSTGSSLR--FKTSGTKTGF 792
OY 79 --TDNITFLGNHSLTFGIDAGTHAGAASATTANKNLTFSGFSLSPDSSSTVTTGQ 136
DB 793 STEKDLITNATGNTLLQVE-GTDMIGKGIYAKKNITFEGGN-ITFESRAVTEIEGN 850
OY 137 GTLSSAGVNL-----ENIRKLVAAGNFSTADGAIKGASFLTGTSGDALEFNNSSSTK 191
DB 851 VTINNANVTLLIGSDFDNHOKPL-----TIKKDVIINSGNLTA 888
OY 192 GGAITTAGARIANNNGYVRFSLNASTSGAIDDEGTSILS-----NNKFY 239
DB 889 GGNIVINAGNLTVESNANFKAITNFTFNVGGLFDKNGNSNISIAKGARFKDIDNSKNLS 948
OY 240 FEGNAKTGGAICMCTKASGSPDLIISNNKTLIFASNVAFETSGAIIHAKK--LALSOGF 297
DB 949 ITTNSSTYRTIISGNTINKNDLITNE-----GSDTEMQIGDVSQKEGNLTJSSDKI 1003
OY 298 TEFLNNVSSATPKGA-----ISIDASGELSLAETGNITFVRN----- 337
DB 1004 -----NITKOITIKAGVDGENDSDATNNANLITKTELKLPQDNLINISGFNKAETAKDG 1058
OY 338 ---TLTTGSDTTPKRNAINISNGK---FTELR-----AAKHHTFFPDPTISEGSSDV 387
DB 1059 SDLTIGNTNSAD-----GTNAKVTYFNQVKDSKISADHKYTLHKSKEVETSSNNNT 1109
OY 388 LKINNGSAGALNRYOGTILFSGETFLADELKYADNLKSFOTPOVSLSGKLLQKVTLE 447
DB 1110 EDDSDNNAGL-----TIDAKNVTYNNNITSHKAVISATISEITTKTGTTIN 1156
OY 448 STS-----FSQENGSLG---MDSGT-TLSTAGSITITNLG-----INVDISGLKOPVSL 494
DB 1157 ATTGAVEITAGQTSILGIESSSGVTLLATGALAVSNISGNFTVTVANSALTTLAGS 1216
OY 495 TAKGASNKVIVSGKLNLDIEGNIYESHMFSDQLFSLIKITVDADVDPNDISSLIPIV 554
DB 1217 TIKG-TESTVTSQSG--DIGGTISG-----TVEVKATESLTTQNSKTIK 1259
OY 555 AEDPNSYGFQOGAMVNMWTTDT-----ATNTKEATA 585
DB 1260 AT-----TGEANVTSATGTIGGTISGNTVNVTA 1287

```

```

Db 1299 ISGTVITTDASGRLTIV--GST-INGNTSVTTSOSGDIETISGNTVAVTASGDLT 1355
QY 663 IGSA 667
Db 1356 IGSA 1360

RESULT 7
US-08-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-038-682-2

Query Match 5.08; Score 238; DB 1; Length 1536;
Best Local Similarity 21.18; Pred. No. 2.6e-09;
Matches 134; Conservative 99; Mismatches 235; Indels 168; Gaps 27;

QY 34 SFNENID-----SGTFPKTSATYSLTGDFV---FYEPEGKPLPSGCFKOT----- 78
Db 736 SFNENISVSGSGSDFLLASSSVQTPGVYINSKYINVTGSSLR---FKTSGSTGTG 792
QY 79 --TDNLFLNGHSLRFGFIDAGTHAGAASSTANKMLTFSGFSLSFSDSPSTTTTQ 136
Db 793 STEKDLFLNATGNITLQVE-GTDMIGKGIYAKKNITFEGGN-TFSGRKAVTEIEGN 850
QY 137 GTLSAGGVN-----NIRKIVYAGNFTADGAIKGAFLTLTGSGDALFSNNSSTK 191
Db 851 VTINNANVTLLGSDPNDHOKPL-----TIKKDYIINSGMLTA 888
QY 192 GGAATATAGARIANTGCVPLSNIASTSGALIDEGSTILS-----NNRFLY 239
Db 889 GGNIVNLAGNLTVESNMFALITFTFNVGGLPDKNGSNISIAKGARRKIDNSKNLS 948
QY 240 FEGNAATGTGATCNTRKASGPELILSNKTLIFASVAVETSGGAIHAKR--LALSSGGF 297

```

```

Db 949 ITNWSSTYRTIISGNTNNKNDLINTNE-----GSDPEMOIGGDVSOQKGNLTISSDKI 1003
QY 298 TEFIRANNVSSATPKGCA-----ISIDAGELSLSAEFGNITFVRN----- 337
Db 1004 -----NITKOITTAGVDGSDSDATNNANLTIKTELKLTODLINSFNKAEITKKG 1058
QY 338 ---TLTTGSDTPPKRAINIGSNK---FTELR-----AAKNHTIFEPDPTSGTSSDV 387
Db 1059 SDLTIGNTNAD-----GTNAKVTFENQVADSKISADGHVTLHSKVTSGSNNT 1109
QY 388 LKINSGAGALNPIYQGTILPGLTDELKADNLKSSFTOPVLSGKILLQKVTLE 447
Db 1110 EDDSDNNAGL-----TIDAKNVTYNNNITSHAASVISATSGETTTKTGTIN 1156
QY 448 STS-----FSQEAGSLG---MDSGT-TLSTTAGSITLTNG-----INVDLIGKQVSL 494
Db 1157 ATGTVNVEITQOTGSLIGLIESSGCVTLTNEGALAVSNISGNTVTVTANSGALTTAGS 1216
QY 495 TAKGASNKVIVSGKLNIDIEGNIYESHMFSDQLFSLKITVDADVDTNVDISLIPV 554
Db 1217 TIKG-TESVTTSSQSG--DIGGTISG-----TVEVKATESLITQSNKIK 1259
QY 555 AEDPENSEYFQGGWNVMTDT-----ATNTEKATA 585
Db 1260 AT-----TGEANVTSATGTIGTISGNTVAVTA 1287

RESULT 8
US-08-302-832-2
; Sequence 2, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```



```

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0810
INFORMATION FOR SEQ. ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

```

```

Query Match 5.2%; Score 247.5; DB 2; Length 1529;
Best Local Similarity 22.0%; Pred. No. 4.9e-10;
Matches 162; Conservative 99; Mismatches 272; Indels 205; Gaps 35;

```

```

QY 11 SSVAFSCHLQSLANEELSPD-SFNGNIDSGTFPTKTSATYSLGDFEYEPGKGP 69
DB 677 SSVAFSCHLQSLANEELSPD-SFNGNIDSGTFPTKTSATYSLGDFEYEPGKGP 69
QY 70 LSDCFKQOTD-----NLTEFLNGHSLTFGFDAGTHAGAASTANKNL 114
DB 732 -SNFSLQOTDSFNEYSKHAINSHMLTILG--NVTLG-----GENSSSTTGMINI 782
QY 115 TFSGFSLSDDSPST---VTTGGTSSAGAVNLENIRKLVAGNFTADGAIKGA 170
DB 783 TNKANVTLOADTSNMTGKRLTLGNSVEGNLSLTGANANIV-GNLSIAEDSTFKGE 841
QY 171 ---SFLTGTSGDLFESNNSSTGGAIAITAGARINANTGYVFLSNIASTSGALDDE 227
DB 842 ASDMLNTGT---FTNNGT-----ANINIKGGVVLQGD----- 873
QY 228 GTSLSNNKFLYEGNAKTTGCAICNTKASGPELIISSNKT-----L 271
DB 874 -----NNK-----GCLNITTNASGDTQTLINGNITNEKCDLNIKIKADAET 915
QY 272 IFASNVATSGGAIAHAKKALLASGGEFTFLRNVSATPRGAIISIDASGELSIAETGN 331
DB 916 QIGENISQKEG-----NLFISSDKV-----NITNQTIRAG---VGGGRSDSEAEANAN 961
QY 332 IT-----FVRNLTITGSGTDTPKRNAINIGSNK---FTELRAAKHNTI 372
DB 962 LTTQTKELKLAGDLINISGFKAEITTAARNGSDLTIGNASGNAADKRYTFPKVXSK--- 1017
QY 373 FFYDPTISEG---TSSDVLKINNSAGALNLPYQGTILFSGEFLTFADELKVAADNLKSSEFTQ 429
DB 1018 -----ISTDGHNTVLTNESEVTSNNGSSNAGNDNS-----TGLTISAKIVTYANNVTSKHTI 1067
QY 430 PVSLSGKLLQKGVLTLESFSGOEA---GSLIG--MDSGTLSTAGSITTTNIGLINV 483
DB 1068 NISAAAGNVTTKEGTTINATGTSVEYTAONGTIGKNTITSONVTVATENLVTFENAVINA 1127
QY 484 DSLQKQPVSLTAAAGANK---VIVSGKMLIDIEGNIYSHMSHQOLSILKIT--VD 538
DB 1128 TS---GTVNISTGTGIDKIGIESTSGNVN-ITASGN-----TLKVSNIITG 1168

```

```

QY 539 ADVDFNVDISLIPVPAEDPNSSEYFGOGMNVNMTTDTA-----TNTKEATATWTKGFV 593
DB 1169 ODVTATADAGALTFTAGSTISATG-----NANITTTGIDNGKVESSSGSVTLVATG-- 1221
QY 594 PSPEKSAALVCNTLWG---VFTDIRSLQOLVEIGATGMEHKOGFWWSMTNPLHKTGDE 649
DB 1222 -----ATLAVGNISGNTVITADSGKLTSTV--GST-INGTNSVTTISQSGDIEGTISG 1272
QY 650 NRKGFRTSGGYVIGGSA 667
DB 1273 NTVNVTASTGDLTIGNSA 1290

```

```

RESULT 5
US-08-719-641-10
Sequence 10, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Matlare, Ltd.
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0810
INFORMATION FOR SEQ. ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

```

```

Query Match 5.2%; Score 247.5; DB 4; Length 1529;
Best Local Similarity 22.0%; Pred. No. 4.9e-10;
Matches 162; Conservative 99; Mismatches 272; Indels 205; Gaps 35;

```

```

QY 11 SSVAFSCHLQSLANEELSPD-SFNGNIDSGTFPTKTSATYSLGDFEYEPGKGP 69
DB 677 SSVAFSCHLQSLANEELSPD-SFNGNIDSGTFPTKTSATYSLGDFEYEPGKGP 69

```

DB 875 TINISAAAGNVTTKGCTTINATGTSVEVTAQNGTICKNGTTSQNVTVTAFTENVTENAVI 934  
QY 482 NVDSLGLKOPVSLTAKGASNK---VIVSGKLNLDIEGNIYESHMFSDHOLFSLKIT-- 536  
DB 935 NATS---GTAVNISRTKGDIKGIESTSGNVN-ITASGN-----TLKASNI 975  
QY 537 VDADVDTNVDISLLIPVAEDPNSEYFGOGMNVMTDTLA-----TNTKATATWTKTG 591  
DB 976 TGDVTVTADAGALTGTAGSTISATTG-----NANITTKTGDKNGKVESSSGCVTLVATG 1030  
QY 592 FVSPERSKALVCNTLMG---VFTDIRSLDQLVIGATGMEHKOGFWVSSMTNPLHKTG 647  
DB 1031 -----ATLAVNGISGVTYVITADSGKLSTV--GST-INGTNSVTTSSGSGDIEGTI 1079  
QY 648 DENRKGFRHTSGGYVIGSA 667  
DB 1080 SGNVTNVNTASTGDLITGNLSA 1099  
RESULT 3  
US-08-617-697-9  
Sequence 9, Application US/08617697  
Patent No. 5977336  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESS: Shoemaker and Maltare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/617,697  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 05-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Beikstesser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-557  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1599 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-617-697-9  
Query Match 5.2%; Score 250; DB 2; Length 1599;  
Best Local Similarity 22.3%; Pred. No. 3.4e-10;  
Matches 165; Conservative 103; Mismatches 284; Indels 188; Gaps 36;  
QY 11 SSVLAFSCHQSLANBELSLPDDSDNSGNDISGTFPKTSATYSLTGDFV-----FYERG- 65

DB 725 ASIMF-----KSNNAVAFENEDISVSG- GSVNFEKLNSSSSMIOFPGYI IKSQNNVSGG 779  
QY 66 -----KGTPILSDSCERQOTDNLTPFGNGSLTFEGTIDAGTAGAASATTKANKLTFSGFS 120  
DB 780 STLNLKAEGETFAFISEND-LNLNATGNTITRQVE-GTDSRVNKKGYAAKKNITFEKGN 837  
QY 121 LSPDSSPST-----VTTGGCTSSAGVNL-ENIRKLVAANGNSTADGAIKASFL 174  
DB 838 -TTFGSOKATTEIKCNVNTINKNTNATLRGANFAEKSPLNAGN----- 880  
QY 175 TGTSDALFSSNNSSTKGAATATTAGARIANTGYRFLSNASTSGAIDEGTSLSN 234  
DB 881 -----VINNGNLTTAGSINIINGNLVSKGANLQATNTNFTNVASFPDNGAS----- 928  
QY 235 NKELYEGNAKTTGCA-----ICNTKAGSPELITSNKT--LIPASNAETS----- 281  
DB 929 -----NISIRGAGAKFKDINTNSLN---ITNSDTYRTYRIKGNISNKSGLNII 976  
QY 282 -----GGAHAKK--LALSOGFTFLRNVSATPKGASISDAGSLSAET 329  
DB 977 DKSDAEIQIGNISOKEGNLTSSDKV-----NTNOITIKAG--VEGGRSDSEAEIN 1028  
QY 330 GNIT-----FVRNLTFTTGSTDTFPRNAINISNGK--FTELRAKNH 370  
DB 1029 ANLTQTEKELKAGDLNISGFNKAETAKNGSDLTIGNASGNMADAKVTPDKVDSK-- 1086  
QY 371 TTFYDPITSEG--TSSDVLKINNGSAGALNPYGTILFSGETLTADELAVADNLKSSF 427  
DB 1087 -----ISTDGHVNTLNSSEVTSNGSSNAGNDNS-----TGLTISAKDVYNNVNTSHK 1134  
QY 428 TQPVLSGKLLQKGVLTLESTFSQEA-----GSLG--MDSGTLSTTAGSITTNIGI 481  
DB 1135 TINISAAAGNVTTKGCTTINATGTSVEVTAQNGTICKNGTTSQNVTVTAFTENAVI 1194  
QY 482 NVDSLGLKOPVSLTAKGASNK---VIVSGKLNLDIEGNIYESHMFSDHOLFSLKIT-- 536  
DB 1195 NATS---GTAVNISRTKGDIKGIESTSGNVN-ITASGN-----TLKASNI 1235  
QY 537 VDADVDTNVDISLLIPVAEDPNSEYFGOGMNVMTDTLA-----TNTKATATWTKTG 591  
DB 1236 TGDVTVTADAGALTGTAGSTISATTG-----NANITTKTGDKNGKVESSSGCVTLVATG 1290  
QY 592 FVSPERSKALVCNTLMG---VFTDIRSLDQLVIGATGMEHKOGFWVSSMTNPLHKTG 647  
DB 1291 -----ATLAVNGISGVTYVITADSGKLSTV--GST-INGTNSVTTSSGSGDIEGTI 1339  
QY 648 DENRKGFRHTSGGYVIGSA 667  
DB 1340 SGNVTNVNTASTGDLITGNLSA 1359  
RESULT 4  
US-08-728-470-10  
Sequence 10, Application US/08728470  
Patent No. 5928651  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESS: Shoemaker and Maltare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2001, 01:27:18 ; Search time 52.09 Seconds  
(without alignments)  
366.824 Million cell updates/sec

Title: US-09-446-677B-2

Perfect score: 4782

Sequence: 1 MKTSPWLVLSVLAFSCHL.....MELRGSSRNVDVGTKLRF 928

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents-AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	250	5.2	1338	2 US-08-728-470-9	Sequence 9, Appl1
2	250	5.2	1338	4 US-08-719-641-9	Sequence 9, Appl1
3	250	5.2	1599	2 US-08-617-697-9	Sequence 9, Appl1
4	247.5	5.2	1529	2 US-08-728-470-10	Sequence 10, Appl1
5	247.5	5.2	1529	4 US-08-719-641-10	Sequence 10, Appl1
6	244.5	5.1	1600	2 US-08-617-687-10	Sequence 10, Appl1
7	238	5.0	1536	1 US-08-038-682-2	Sequence 2, Appl1
8	238	5.0	1536	1 US-08-302-832-2	Sequence 2, Appl1
9	238	5.0	1536	2 US-08-530-198-2	Sequence 2, Appl1
10	238	5.0	1536	2 US-08-469-880-2	Sequence 2, Appl1
11	238	5.0	1536	2 US-08-728-470-2	Sequence 2, Appl1
12	238	5.0	1536	2 US-08-617-697-2	Sequence 2, Appl1
13	238	5.0	1536	2 US-08-719-641-2	Sequence 2, Appl1
14	227.5	4.8	1026	2 US-08-614-377A-7	Sequence 7, Appl1
15	227.5	4.8	1026	4 US-09-142-648B-7	Sequence 7, Appl1
16	226.5	4.7	1160	1 US-08-194-290-7	Sequence 7, Appl1
17	223.5	4.7	1160	1 US-08-808-559A-24	Sequence 24, Appl1
18	211	4.4	1612	1 US-08-169-927-2	Sequence 2, Appl1
19	192	4.0	1477	1 US-08-038-682-4	Sequence 4, Appl1
20	192	4.0	1477	1 US-08-302-832-4	Sequence 4, Appl1
21	192	4.0	1477	2 US-08-530-198-4	Sequence 4, Appl1
22	192	4.0	1477	2 US-08-469-880-4	Sequence 4, Appl1
23	192	4.0	1477	2 US-08-728-470-4	Sequence 4, Appl1
24	192	4.0	1477	2 US-08-617-697-4	Sequence 4, Appl1
25	192	4.0	1477	2 US-08-719-641-4	Sequence 4, Appl1
26	177	3.7	2123	4 US-08-968-685A-10	Sequence 10, Appl1
27	175	3.7	674	1 US-08-317-522A-3	Sequence 3, Appl1

28	175	3.7	674	1 US-08-439-818A-3	Sequence 3, Appl1
29	175	3.7	674	2 US-08-751-965-3	Sequence 3, Appl1
30	175	3.7	674	2 US-08-738-975-3	Sequence 3, Appl1
31	175	3.7	674	2 US-08-728-626-3	Sequence 3, Appl1
32	175	3.7	674	3 US-08-808-599A-3	Sequence 3, Appl1
33	174	3.6	749	1 US-08-317-522A-2	Sequence 2, Appl1
34	174	3.6	749	1 US-08-439-818A-2	Sequence 2, Appl1
35	174	3.6	749	2 US-08-751-965-2	Sequence 2, Appl1
36	174	3.6	749	2 US-08-738-975-2	Sequence 2, Appl1
37	174	3.6	749	2 US-08-728-626-2	Sequence 2, Appl1
38	174	3.6	749	3 US-08-808-599A-2	Sequence 2, Appl1
39	171	3.6	2353	4 US-09-377-155-33	Sequence 33, Appl1
40	171	3.6	2353	4 US-08-913-942-4	Sequence 4, Appl1
41	166.5	3.5	679	4 US-08-913-942-15	Sequence 15, Appl1
42	165	3.5	1912	1 US-08-409-995-4	Sequence 4, Appl1
43	165	3.5	1912	3 US-08-685-467-4	Sequence 4, Appl1
44	161.5	3.4	666	2 US-08-737-716-14	Sequence 14, Appl1
45	161.5	3.4	1848	4 US-08-296-791-6	Sequence 6, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-728-470-9  
; Sequence 9, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728,470  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-633  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1338 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-728-470-9

**THIS PAGE BLANK (USPTO)**

OY	397	----	ALNPGCTTISEETLTAAELKVAADLKSSFTQPSLSGGKLLKQGYT--	LEST	449
Db	357	VANPDVLNAGGCVNLKAINITTTD-----	-----	ANGVSVFAGAGTINYSGT	399
OY	450	SFSQDAG--SLLGMDSGITL-----	-----	STAGSITITNLGIN-----	VDLSLGLK 489
Db	400	VGGQGGKNFNTVALENGTTFVXFKLGATNFNGNTTFAASTLQIGNATYADCVASADGTGIY	459		
OY	490	Q-----PVSLLAK-----	GASNKVIYSGKLNLDIE-----	GNIVSHMFSDOLF	530
Db	460	EFVNTGPTITVLNKOAPAVNLKQITVSGRGVAVINISGAGNHGCAVDTDTAFENSSLSG	519		
OY	531	SL-----	LKTIYADVDVNDISLLPVPAE-----	DPNSE	561
Db	520	AVVFLPRIPNDAGNTMPLTIKTYCNKTKAKGDVPSVYVLGVDVIADGVYIGQNNI	579		
OY	562	YGF-OGQNN--VNMVT-----	DTATNTEKATATWTGTGFVPSPERKSALVCTMLNGVFT	612	
Db	580	VGLGSGDNGIIVATATLYAGISTLANNO--	GTVLTSGGVPTNP-----	GTVYGLST	629
OY	613	DIRSLQQLVEIGANGMEHKQGFWMSSMTNF--	LHKTDENKRPFRHSGYV-----	662	
Db	630	G-----	IGAS--FKQVTFPTDYNNGIINLIATATATINDGVYVTTGSGIAGIGFDGKI	678	
OY	663	-----	IGGSAHTPKDQDLTFEAFCHLFARDKDCFIANNHNSFTVGGTLEFFKHSHTLQONPYLR	718	
Db	679	TLGSVNGNGNRPADGLISNSTSMIGTK-----	ANNQTVIYLLGNAF-----	720	
OY	719	LGRAKFSASAIEKPRELIPALADVQVSFSSDN--	RMEHTYTLPSSEGSW-----	SNE	770
Db	721	VGNIDSDTPV-----	ASVRFETGSDSAGLQGNIVSOVIDFTGYNLGIIVNSNI	768	
OY	771	CIAGIGIGLDLPEVLSPNPHLEKFTPIPMKTEMYVYVSONSFESSDORGSGISGRLLNLSI	830		
Db	769	ILGG-----	TTAINGKIDLYNTLITLFSAGSTMGKNNTSITITTLTA-	810	
OY	831	PVGAKFVQGDIG-----	DSYTYDLSGFVSDVYRNNPOSTATLVMSPDSMKIRG-	879	
Db	811	-----	NGNIGHIIVILEGAGVNTTTGTGTTIKQVDNANANFGTQTYTL-----	IQGG	857
OY	880	-----	GNLSROAFLLRGSNNYYNSNCELFGHYAMELRGSSRNT	918	
Db	858	ARENGTLGSPNFVATGNSRF-----	NYSL-ITAAQDY	890	

Search completed: October 2, 2001, 03:31:52  
Job time: 227 sec

```

Db 689 --AAGTNLG-----SATNPLAEINFGSGVNVDTVLNVGSGVNLATNTTTDANVGSFV 741
Oy 497 --KGASNKVIVSG-----KLNLDIEGNIYESHMFSDQLFSLKITVDADVDPTNV 546
Db 742 FMAGCTN--IVSGVGGGQGNKNTVALENG-----TYKFLGNATFNQNTT 786
Oy 547 ISSLIIPVADPNSEKFGGQGMVNVMTTD--TAINTKATATWTKTGFPSPERKSALVCN 605
Db 787 IAA-----NSTLQIGS-----NTADCVASADGTGIVEFVNTGPIVTLNKAQAPVN 833
Oy 606 TLMGVFTDIRSLOOLVEIGATGMEH-----KGFVSVSMTNPLHK 645
Db 834 ALKQIVVSGPNVYINELGAGNHGAVTDTIAFENSLSGAVVLPKGIPTFNAGMTMPL 893
Oy 646 T-----GDERNKGFRTS-----GGYVIGSAHPKDDLFACCHLFARDKDC 689
Db 894 TISSTYGNKAKAFDPSPVYVLGVDSIADGQVIGDNNIVGGL-----GSDN 942
Oy 690 FIANNSTRYGGTLFFKHS--TLQ-----PONTLRL--GRAKFSSEAIKEPREI 736
Db 943 GIIVNATTLVYAGISTLNNGGVTLSGVPNTPGTVYGLGTGIGASKFK----- 991
Oy 737 PLALDVGVSHSDNRM-ETHYTSLPESGSMNECAGIGLD---LFPVLSNPHLPK 791
Db 992 -----QVFTTYYNMLGNIAATNATINDGVTVTGGIAGIGFDGKITLGSVNGNVRF 1045
Oy 792 KTEIPOMKVEVYVSONSFESSSDGKFGSIRGLNLNIPVGA-KFYOGDIG----- 842
Db 1046 ADGILSNTSMICTYKANNTVYTLGNAF-VGNIGSDPTVAVSRPTGSGAGLOGNIT 1104
Oy 843 ----DSYVYDLSGFVSVYRNN-----POSTATLVNSPDSWKIRGNLSRQAF 887
Db 1105 SOVIDGCTYML-GIVNSNITLIGGTTAINGKIDLVNTLTFASGTSFW---GNNTSIETT 1160
Oy 888 LRGSNVYVNSNCELFGHAMELRGSSRN 917
Db 1161 LTLANGN-----IGHIVT-LEGAQVN 1180

RESULT 15
120K RICRI STANDARD; PRT; 1300 AA.
AC P14914;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE 120 KDA SURFACE-EXPOSED PROTEIN.
GN P120.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxId=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R.
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii.;
RL Mol. Microbiol. 3:1579-1586(1989).
CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS
CC RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -1- DISEASE: RICKETTSIA RICKETTSII IS THE CAUSATIVE AGENT OF ROCKY
CC MOUNTAIN SPOTTED FEVER (RMSF).
CC -1- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPITOPES
CC CONFERRING ANTIGENICITY TO THE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: X16353; CAA34402.1; -.
DR PIR: S07575; S07575.
KW Antigen; Glycoprotein; S-layer.
FT CARBOHYD 66 66 N-LINKED (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (POTENTIAL).
FT CARBOHYD 436 436 N-LINKED (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (POTENTIAL).
FT CARBOHYD 799 799 N-LINKED (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (POTENTIAL).
FT CARBOHYD 826 826 N-LINKED (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (POTENTIAL).
FT CARBOHYD 861 861 N-LINKED (POTENTIAL).
FT CARBOHYD 879 879 N-LINKED (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (POTENTIAL).
FT CARBOHYD 1116 1116 N-LINKED (POTENTIAL).
FT CARBOHYD 1128 1128 N-LINKED (POTENTIAL).
FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
FT CARBOHYD 1146 1146 N-LINKED (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (POTENTIAL).
SQ SEQUENCE 1300 AA; 132801 MW; E09E52C3B647243D CRC64;

Query Match 4.1%; Score 196; DB 1; Length 1300;
Best Local Similarity 19.8%; Pred. No. 0.0012;
Matches 210; Conservative 142; Mismatches 345; Indels 366; Gaps 53;

Oy 21 QSLANEELSPDPSFNGNDISGTF-----TPKTSATYTSITGVFFPEPKGPIPLSDSCF 75
Db 29 KTAASKVTTIYQDSNF-GMTDFGNIAAQIKVPAITLITGNFTGDA--SNP----- 75
Oy 76 KQPTDNLTFNGCHSLTFGIDAGTHGAAASTANKNLT---FSGFSLSPDSSPTTV 132
Db 76 -NTAGVITFPANG-----TLESASADANVAVNRNITRAIEASGAGVQLSGTHAAEL 125
Oy 133 TTGQGTLLSAGVNLNIRKLVVAGNFTADGAING---ASFLLTGTSGDALPSNNS 189
Db 126 RLG-----NAGSI-----FKLADGVYINGKVQTLVYGALAACTITLDGSA 167
Oy 190 TKGAIATITGA-----RIANMTGVYREL--SNIASTSGAID---DEGTSLT---NN 235
Db 168 TITGDIGNAGGAALALORTITLANDAKKTLLVIGAGGTTIDLOANGSTIKLTSTQN 227
Oy 236 KFLYFEGNAKTTGGAICNTKASGPELLI-----SNKTLIFASNAVETSGAIIA 287
Db 228 IYVDFDLAATDGTGVVDSASLTNAQTLITNGKILGITGANNKLT-----GQPNIGS 278
Oy 288 KKLALSSGGEFTFLRNNVSSAIPKGAISIDASGELSLSAETGNITFVRNTLTGSTD 347
Db 279 SKTVLSNG-----NV-----AINELVIGND-GAVOFADHTVYLITRTT--- 314
Oy 348 PKRNAINIGSNKGFTELRAKANTIFFYDPTISEG-----SSDYKLKINNGSAG 396
Db 315 -----NAAGQK-----IIFNPVYVNGTTLAAAGTLGSAIPLAEINFGSGK 356

```



```

CC -----
DR EMBL: Z15035; CAAT8753.1; -
DR EMBL: X75780; CA53406.1; -
DR EMBL: Z28068; CA81905.1; -
DR PIR: B44402; B44402.
DR PIR: S39173; S39173.
DR PIR: S44518; S44518.
DR SGD: S0001551; NUP100.
KM Nuclear protein; Transport; Repeat.
FT DOMAIN 33 571 G-L-F-G.
SQ SEQUENCE 959 AA; 99988 MW; D3985F9901BBA51 CRC64;

Query Match 4.2%; Score 203; DB 1; Length 959;
Best Local Similarity 21.1%; Pred. No. 0.00032;
Matches 191; Conservative 107; Mismatches 331; Indels 278; Gaps 40;

QY 12 SVLAFCSHQSLANELLSPDSE--NGNIDS-----GTFPTKATATYSLTGDV 59
DB 12 SNLSFGSNTSFGGQSQSPNSLFGNSNNNNSTNNAGSGFGFTSAGSNSLSFGN- 70
QY 60 FFEYPCGCTPLSDSCFEQOTDNLTF---LNGHSLTFCFIDAGTHAGAASTANKULTF 116
DB 71 -----NNTQNNCAFGGSMGATONSPPGSLN-----SSNASNGTFF 105
QY 117 SGFSL-SPDSSPSTVYTGOGTSLSAGVNLNIRKLYVA--GNFSTADGAIKGSFL 173
DB 106 GGSSSMGSGFGNTNNANNNNSNSNSPFGNKPRTGTLFGSONNSAGTSSLFEGOSTS 165
QY 174 LVTGSGDALESNNSSSTKGAATAGARIANNMGVYRFLSNIASTGAIIDEGT----- 229
DB 166 TTGTFGNTGSSFGTGLNGNSNIFGAGNNSQNSMTGSLF--GNOQSSAFGNNQGGSLFGQ 224
QY 230 -SILSNKPKLFEBNNAKKTGCAICNTKASGPELIISSNKKTLIFASNVAETSGAIHAK 288
DB 225 QSONTNNAF---GN-ONQAGSFGSKPVGSGSLFGQSNNTLGNNTNNGRLFGQNSS 279
QY 289 KIALSSGSGFTEFLRNYSATP-----KGAIAIDASGEL--S 324
DB 280 NQSSNSGL--FGONSNSSTQGVFGONNNOMQINGNNNSLFGKANTFSSNSGGLFGQ 337
QY 325 LSAETGNTFVRNLTLTGTTGTDTPKRNAINIGSNGKFTLEAAKNHTIFPYDPTISGTS 384
DB 338 NNOQGGGGLFGQNSQTS-----GSSGGLFGQNNQKQPN-----FTQGNTC 377
QY 385 SDVIAKIN-----GSAGALNPYQGITLFSGETLT-ADELAKVADNKKSFQPV--SL 433
DB 378 IGLFGONNNOQOSTGLFGAKPAGTGSLLFGGNSSTQPSNLFGTNTVPTSNTOQSGNSL 437
QY 434 SGGKLLQ-----KGVTLSTSFQDAGSLGMD--SGTTLSTTAG----- 472
DB 438 FGATKLTNMFPGNPNTANQSGSGNSLFGTKPASTGSLFGNNTASTTVPSTNGLFGNNAN 497
QY 473 -SITTNMGI-----NVDISLGLKOP----- 491
DB 498 NSTSTNTGTFGAKPDQSKPALGGGLFGNSNSNSTIGQNKPVFGGTQNTGLFGATGT 557
QY 492 ----VSILTA--GASKNVIYSGKLNLDIGCNYESHM-----FSDQFLSL 533
DB 558 NSSAVGSGTGLFGONNNTLVGQNPVNNNTQNALGLTAAVPSLQOAEVTEQLFSKI 617
QY 534 KI-----TYDADVDQTNVDISS--LIPVP--AEDPNSEKFGQGNVNMVTTD 575
DB 618 SIPKSTINPVKATTSKYNNADMKRNSLITSAYRLAPKPLPAPSSNGDKKFKGKGTLEKSD 677
QY 576 TATNTKATATWTKGTFVSPERKKSALVCTLWGVF-TDIRSLQOVLIELATGMEHKQGF 634
DB 678 RGSSTNS-----ITDPE--SSYLSNNDL--LFDPRRRLKHLV-----IKNNKNL 719
QY 635 WVSMTNFKLKTGDENK-----GFRHNSGGVYIGGS-----AHTPKDGLTFEAF 679
DB 720 -----NVINHNDKASKVLVYFTTESASKDOASSSIAASKLTERKAHSPQDL----- 768

```

```

QY 680 CHLEARDCEFIANNNSRTYGTLEFKHSHTLQPONYRLGRAKFSESAIEKPREIPLA 739
DB 769 -----KD-----DHDES-----TPDPQSPSPGSGTISIPMIENEKISSKYPGL 805
QY 740 LDYQVSF 746
DB 806 LNDVTF 812

RESULT 13
MSB2_YEAST STANDARD; PRT; 1306 AA.
AC P32334;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MSB2 PROTEIN (MULTICOPY SUPPRESSION OF A BUDDING DEFECT 2).
GN MSB2 OR YGR014W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92383951; PubMed=1514328;
RA Bender A., Pringle J.R.;
RT "A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence
RT defect."
RL Yeast 8:315-323(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=528BC;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT chromosome VII."
RL Yeast 13:1077-1090(1997).
CC 1- P1M: O-GLYCOSYLATED IN THE SER/THR-RICH REGIONS (PROBABLE).
CC -1- SIMILARITY: SOME, TO YEAST HKR1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77354; AAA34798.1; -
DR EMBL: Z72799; CA96997.1; -
DR PIR: S25370; S25370.
DR SGD: S0003246; MSB2.
KW Transmembrane; Glycoprotein; Repeat.
FT DOMAIN 698 816 7 X 17 AA TANDEM REPEATS.
FT REPEAT 698 714 1.
FT REPEAT 715 731 2.
FT REPEAT 732 748 3.
FT REPEAT 749 765 4.
FT REPEAT 766 782 5.
FT REPEAT 783 799 6.
FT REPEAT 800 816 7.
SQ SEQUENCE 1306 AA; 133114 MW; 67D5D984D5CA4A6D CRC64;

Query Match 4.2%; Score 202; DB 1; Length 1306;
Best Local Similarity 18.7%; Pred. No. 0.00057;
Matches 128; Conservative 135; Mismatches 277; Indels 144; Gaps 23;

QY 4 SIPWLVSSVLARSCHQSLANELLSPDSE-----NGNIDSGTFF--PK 47
DB 344 SAEFLDTSTNSPS-----IVSPSVFVPSQSSSDVASSSTANVSSFSFSDIPQ 393

```

"Evidence for proteolytic cleavage of the 120-kilodalton outer membrane protein of rickettsiae: identification of an avirulent mutant deficient in processing.".

RL Infect. Immun. 60:159-165(1992).

CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOPEDANT DURING INFECTION.

CC -1- FUNCTION: THE 32 KDA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.

CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M37647; AAA26390.1; ALT\_INIT.

DR EMBL: AF161079; AAD42234.1; -.

DR EMBL: AJ235273; CA15140.1; -.

KW Antigen; S-layer; Cell wall.

FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.

FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.

FT VARIANT 257 257 V -> A (IN STRAIN BREINT).

FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINT).

FT VARIANT 1450 1450 A -> S (IN STRAIN BREINT).

FT CONFLICT 178 179 AA -> VC (IN REF. 1).

FT CONFLICT 191 201 TTPAPPLTGA -> INSRSSSYLVS (IN REF. 1).

FT CONFLICT 212 212 T -> I (IN REF. 1).

FT CONFLICT 313 313 O -> L (IN REF. 1).

FT CONFLICT 1104 1104 D -> G (IN REF. 2).

FT CONFLICT 1123 1123 T -> S (IN REF. 2).

FT SEQUENCE 1643 AA: 169854 MW: 735FD392E6346CC CRC64:

SO

Query Match 4.38; Score 208; DB 1; Length 1643;

Best Local Similarity 19.58; Pred. No. 0.00036;

Matches 170; Conservative 128; Mismatches 280; Indels 292; Gaps 40;

QY 9 LVSSVLAESCHLOSLANELSPDPSFN---GNISGTFPKTSATYSLGDFEYEP 64

DB 370 IVDVGLGTTNFKFADSKITITENSGFNGLDQIVAPDFILKGNFGIDV---K 425

QY 65 GKTFPLSDSCKQTDNLTFLNGHSLTFPFLDAGTHAGAAASTAANKNLFPSGLISF 124

DB 426 NNG---NTAGVITTFNAG-----ALVSA 445

QY 125 DSSPSTVVT---GQGLSSAGVLENIRKLVAGN---ESTADGAIK--- 169

DB 446 STDPRIAVTNINAIYEBAGAVY-ELSGHIAELR---LGNGSTFLKADGIVNGPYNQ 500

QY 170 -----ASFLITGSGDA-----LFSNNS-----STK 191

DB 501 NALMNNNALAGSIOLDSDSAITIGDINGVNAALQHTLTLDAKILALDGANITGANV 560

QY 192 GGAI---ATTAGARIAN---NNGYVRFSLNINASTSGAIDDECTSLSNKFLYFEGNAK 246

DB 561 GGAHFPQNGSTIKLTNTONNIIVNFDLITTKGVVD---ASSLTNNQTLTNG--- 613

QY 247 TTGCAICNTKASGPELITSNKTLIFASNAVETSGAIIHAKKLLSSGGFELRNVS 306

DB 614 SIGIVVANTKILA---QNIWSSKTLINAGVA-----INELVTEWNGSVQ-LNHMTY 662

QY 307 SATPKGAISIDAGSELSAETGNITFVRNLTITGSDTP-----KRANINIGS 357

DB 663 LIFK---TINANGQIIVADPLNTNTTLADGTNLGSAENPLTHPATKAANADSLN 719

QY 358 NGKTEFLAANNNHTFPDPTISGTSQVJKNNGS---AGALNIPQG----- 403

DB 720 VGKGVNLVA-----NNTITNDANVGSILHPSGSGTSIVSGVGGQGHKLNNLILDN 770

QY 404 --TILFSGET-----LPADELKAVDN---LKSFTQPVVS----- 432

DB 771 GTTVAFLEDGTTNGTKIEGKSILOISNNYTTDHWESADNKTGLEFVWTDPTVTLNKG 830

QY 433 -----ISG-----GKLLKQVYLTESFSQEA-GSLAMDST-----TL 467

DB 831 AVFGLVQKQIISGPNIIVFNEIGNVGIHGIANSISFENASLSLTPSPGTPADVLTI 890

QY 468 STTASITITNIGT-----NDSIGLKQVSLTKKGSNKVYSGKLNLDIEBNIES 521

DB 891 KSTVANGTVDFENADIVVYSGIDSM-----INNGIIDKKNIILSLGSDNS 938

QY 522 HMFSDOLFSLIKTVADVDPTDNVDSILIPVAPDPSSEFGQGMN-----VNMVT 574

DB 939 ITVNAANTLYSGIRTT--KNNGQTVTLSGMP--NNPGTIYL-GLENGSPRLKQVTFY 992

QY 575 D-----TATNTEATATWTKGFVPSPERKSALVCNTLMG-----VETDIRSLQ 618

DB 993 DYNNGSIANNVTLINDVYTLTGAGIDPDAKITLGSVNGNANVREVDSTFSDPRSMI 1052

QY 619 OLVEIGATGMEKQGFVWSSMTNF-----LKTGDENRKRGRHNSGGVYIGSAHTEPKD 672

DB 1053 VATQANKGTIVYILGNALVSNIGSLDTPVAVSFETGDSGAGIQ-----GNITYSONI 1103

QY 673 DLETFAPCHLFARDKCFIAHNSRYTGT 702

DB 1104 DPGTT-----NLITLNSNVLGGGT 1123

RESULT 12

N100\_YEAST STANDARD: PRT: 959 AA.

ID N100\_YEAST

AC 002629:

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE NUCLEOPORIN NUP100/NSP100 (NUCLEAR PORE PROTEIN NUP100/NSP100).

GN NUP100 OR NSP100 OR YKL068W OR YKL1336.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI\_TaxID:4932;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=93054906; PubMed=1385442;

RA Wente S.R., Rount M.P., Blobel G.;

RT "A new family of yeast nuclear pore complex proteins.";

RL J. Cell Biol. 119:705-723(1992).

RN [2]

RP SEQUENCE FROM N.A.

RA STRAIN=5288C;

RC MEDLINE=94378724; PubMed=8091863;

RX Raamsen S.W.;

RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the NUP100 gene, an open reading frame (ORF) possibly representing a nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in addition to seven ORFs with weak or no significant similarity to known proteins.";

RL Yeast 10:569-574(1994).

CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.

CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.

CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.

CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.

CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



CC -1- SIMILARITY: STRONG, TO BODETELLEA PERTACTIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U36840: AAA79815.1; ALT\_SEQ.  
 DR EMBL: AE000350: AAC73695.1; -  
 DR EMBL: D90889: BAA16514.1; ALT\_INIT.  
 DR EMBL: D90890: BAA16518.1; ALT\_INIT.  
 DR EcoGene: EG31213: YPJA.  
 KM Hypothetical protein; Outer membrane.  
 SQ SEQUENCE 1569 AA; 162774 MW; 0D006A9A32154596 CRC64;

Query Match 4.4%; Score 209; DB 1; Length 1569;  
 Best Local Similarity 20.2%; Pred. No. 0.00029;  
 Matches 197; Conservative 131; Mismatches 356; Indels 290; Gaps 49;  
 QY 40 DSGFTPTKTSA---TYSLTGDFEYEPKGPPL---SDSCFKQYTDMLTFLGNGHSL 91  
 DB 624 NAGSLFVVTGSRAVDITLIINNGKMDYVKDGVFLNAGTQTIVASATSPKANIKGKQT 683  
 QY 92 TFGPTDAGTAGAAASTANKNLTFCGSLISDSSPTVTGQGLSSAG-GVNLNMT 150  
 DB 684 VYGL-----ATEANIE---SGEQIVDGGSTKRTKINGTQTVQVYKGAINDIV 729  
 QY 151 RKLIVAGNFSTADGAIKASFLTT-GTSGDALFSNNS-----STKGA 194  
 DB 730 SGLOQIANTATRESITINGSSQYVNEGGIADENSVLNDGGLDVPREKASATGQSSGAL 789  
 QY 195 IATTAAGART-ANNTGYRF-----LSNASTSGGAIDF----- 227  
 DB 790 VATTTRATRVGTADGVAFEIEQGAANNILLANGVLTVESDTSKTYQVMGGRRTVKT 849  
 QY 228 -----GTSIISNNKFLYFEENNAKT-----GGAICMTKASGSELIISNKT 271  
 DB 850 KATATGTTTLGGEGYIV-EGVANETTINDGGIQTIVSANGAETIKIKIEGGTLYVNDGK- 906  
 QY 272 IFASNAVETSAGAIH-----AKKIALSSGGFTTEELRNVS- 306  
 DB 907 --ANDIYQNSGAALQSTANGIEISGTHQYGTFSISGNLATNMLENGMLVLAGTEAR 964  
 QY 307 -SATPKGAL-----SIDASGELSL-----AET-----GNITFVRNTLT 340  
 DB 965 DSTVKGGAQONLQDSATRVNSGGQYTLGRSKDEFOALARAEDLQVAGGTATVYAGTTLA 1024  
 QY 341 -----TTGTFD--TPKRN-----AINISNGKFT-----ELRAKNHITFE 374  
 DB 1025 DASVSGATGSLMTLPDQNTVPYKLEGAIVRTTSATLTLLGNGVDTLADTLAASRGSVWL 1084  
 QY 375 YDPTSGTSS-----DYKINNG-----SAGALNPYOGTILFSEGLTADLEKADMLKS 425  
 DB 1085 NSNNSCAGTSCNCEKRYVSLNDGDVYLSAQTAAPATTNIIYN-TITTMELSSGNGF-Y 1141  
 QY 426 SFTQPVYSLSGKLLKQGVLTLESTFSQEAAGSLGMDSGTTLSTTAG---STTTNGLIN 482  
 DB 1142 LHRNVASRSDQDLVNNNANAGNFKIFVQDGVSPQSDAMTLVKTGGGDAISFTLGNNGF 1201  
 QY 483 VDSLGLKOPSLAKAKSNVIVSGLNLIDIGNITESHMFHIDQLFSLTKITVADVD 542  
 DB 1202 VD-LGITEYV-LASDGN-----WNLTN-----DVK 1226  
 QY 543 TNVD-ISSLIPVAEDPNSRYFGQGMNVMVTDATNFK-EATATWTKTGFVSPERK- 599  
 DB 1227 PNPDPINPKPDPKDPKPD-----PNPKPDPDPDPDPDP-----PVPERKRI 1267  
 QY 600 ---SALVGN---TLMGVF-IDIRSLDQOLVIGATGMHKGKGFVSSMTNPLHKTGDENRK 652

DB 1268 TPSTAAVLNNAATLPLVDAELNSIRERLNI-MKASPHNNVNGATYNTNNYTDAG-A 1325  
 QY 653 GFRHTSGGYVIGSAAHPRKDLFT-----FACHL-FARDKDCFIAHNSRPT--GGT 702  
 DB 1326 GFETTLTGMYVIGDSRNDIREGITTLGAFMGYSHSHIGFDRG-----GHSSVGSYSLGGY 1380  
 QY 703 LFRKSHHTLOPONYLRIGRAKFSESAIEKPREIPLADVOVSFSHSDNMETHY-TSLP 761  
 DB 1381 ASMHHEGFLDGVYKLRNRKSNVAG-----KMSSGGAANGSYHNGLGHIETGMR 1432  
 QY 762 ESEGSWSNECIAGIGIDLPVLSNPAPLEKTFIPQMKVEMYVYSONSFESSDGGCF- 820  
 DB 1433 FTTGNNMLTPTVASTLG---FTADNPE-----YHLSNGMKRSKYDTRSIY 1473  
 QY 821 -SIGRLNLSIPVG 833  
 DB 1474 RELGATLSTYNNRUG 1487

RESULT 11  
 OMPB\_RICPR STANDARD; PRT; 1643 AA.  
 AC Q53020; Q9ZCM0;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)  
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)  
 DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)  
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].  
 GN OMPB OR SPA OR SPA OR RP704..  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiense; Rickettsia.  
 OX NCBI\_TaxId=782;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-BREINL;  
 RA MEDLINE=91045972; PubMed=2122457;  
 RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;  
 RT "Characterization of the gene encoding the protective paracrystalline-  
 RT surface-layer protein of Rickettsia prowazekii: presence of a  
 RT truncated identical homolog in Rickettsia typhi.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BREINL;  
 RA Moron C.G., Yu X.J., Walker D.H.;  
 RT "Sequence analysis of ompB of Rickettsia prowazekii.";  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RA MEDLINE=99039499; PubMed=9823893;  
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,  
 RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland R.M.,  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria.";  
 RL Nature 396:133-140(1998).  
 RN [4]  
 RP PARTIAL SEQUENCE.  
 RC STRAIN-BREINL;  
 RA MEDLINE=92114896; PubMed=1370573;  
 RA Ching W.M., Carl M., Dasch G.A.;  
 RT "Mapping of monoclonal antibody binding sites on CNBR fragments of  
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia  
 RT prowazekii.";  
 RL Mol. Immunol. 29:95-105(1992).  
 RN [5]  
 RP IDENTIFICATION OF CLEARAGE SITE.  
 RA MEDLINE=92104668; PubMed=1729180;  
 RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;

```

RC STRAIN=WILMINGTON;
RA MEDLINE=94040787; PubMed=8224886;
RX Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT "Cloning and sequence analysis of the gene encoding the crystalline
RL surface layer protein of Rickettsia typhi";
RN Gene 133:129-133(1993).
RP PARTIAL SEQUENCE.
RC STRAIN=WILMINGTON;
RA MEDLINE=92114896; PubMed=1370573;
RX Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNR fragments of
RL the S-layer protein antigens of Rickettsia typhi and Rickettsia
RN prowazekii.";
RM Mol. Immunol. 29:95-105(1992).
RN [3].
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cieplak W., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RL membrane protein of rickettsiae: identification of an avirulent
RT mutant deficient in processing.";
RM Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L04661; AAB48987.1; -.
KW Antigen; S-layer; Transmembrane; Cell wall.
FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1 1354 1645 32 KDA BETA PEPTIDE.
FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
FT CONFLICT 657 657 H -> N (IN REF. 2).
FT CONFLICT 842 842 V -> I (IN REF. 2).
FT CONFLICT 1071 1071 G -> A (IN REF. 2).
FT CONFLICT 1306 1306 G -> S (IN REF. 2).
SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

Query Match 4.4%; Score 211; DB 1; Length 1645;
Best Local Similarity 22.6%; Pred. No. 0.00024;
Matches 177; Conservative 81; Mismatches 223; Indels 302; Gaps 43;

```

```

OY 287 -----AKKA-----LSSGGTEFLRNWSSA 308
DB 676 KIIYAADPIINDTALADGTNIGASAPLSNIHFATKANAGSILHIGGVNIYANNITTT 735
OY 309 TPKGASIDASGELSLSAETG-----NITEVR--NTLTGTTGTPRKRRAI 353
DB 736 DANVGLHPRSGSTISVGTGGGGLKLNILNDGTIVKFLDITNGTKIEGKSIL 795
OY 354 NIGSNGKTEFLRAKN-HTIFFY--DPTT-----SEGTSSDYK-----INN 392
DB 736 QISSVYITDIESADNTGLEFVNPTPTVYLNKGAGVGLKQWVSGPGNIAPNEIGN 855
OY 393 GSAGAL-----NPGYGTIF--SGEFLVDEK--VADLKRSGFOPV-SLSGKIL 439
DB 856 GVAHAIANDSIFENASIGASIFLISGTPLDVLTIRKSYNGKNTVNFAPILIVSGISDM 915
OY 440 LQKG-----VLES-----TSFSQAGSL-----458
DB 916 INNGVIGDQKNIIIALSIGDSNISIVNSNTLYAGIRTKRNGVTLTSGIPNNGTITYG 975
OY 459 IGMDSG-----TTLSTAGSITTNLGINDSLGLKQPVSLTAKGASKYIVSKL 509
DB 976 LGLENGDEKRLQVPTPTTYNNIGSIIATNVTIIND-----VLTGTGIG--IDFDKI 1027
OY 510 NLIDIEGNI-----YESHMSDQLFLKITYDADVDY-----NWDISLIPVPA- 555
DB 1028 TLGSLNGANNAKFDVYRTSHP---TSMIVSTRKANOGTYTIGNALVGNIG--SDIPVASV 1083
OY 556 --EDPNSEYGFQO-----WN-----VNMTPDPAVT--KEAT 584
DB 1084 RPTGSDSVGLQGINHSQINIDFGYNTLINSVDILGGTTAINGEIDLTLNLIIFANGT 1143
OY 585 ATW 587
DB 1144 STW 1146

RESULT 10
ID YPJA_ECOLI STANDARD; PRT: 1569 AA.
AC P52143; P76610; P77017; P77019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHEtical 98.4 KDA PROTEIN IN ALPA-CABD INTERGENIC REGION (P949).
GN YPJA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Horuchi T.,
RA Ikemoto K., Inada T., Isono K., Itoh T., Kanai K.,
RA Kasai H., Kashimoto K., Kim S., Kimura S., Kitagawa M.,
RA Kitakawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H.,
RA Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Tagami H., Takemoto K., Wada C.,
RA Yamamoto Y., Yano M.;
RL Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).

```

```

RESULT 8
AIDA_ECOLI STANDARD; PRT; 1286 AA.
ID AIDA_ECOLI
AC 003155;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE ADHESIN AIDA-I PRECURSOR.
GN AIDA-I.
OS Escherichia coli.
OG Plasmid pIB6.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RC STRAIN=2787 (0126:H27);
RX MEDLINE=9232638; PubMed=1625582;
RA Benz I., Schmidt M.A.;
RT "AIDA-I, the adhesin involved in diffuse adherence of the
RT diarrhoeagenic Escherichia coli strain 2787 (0126:H27), is
RT synthesized via a precursor molecule."
RL Mol. Microbiol. 6:1539-1546(1992).
CC -1- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC TO EPITHELIAL CELLS.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X65022; CAA46156.1; -
DR PIR: S28634; S28634.
KM Cell adhesion; Signal; Outer membrane; Plasmid.
FT SIGNAL 1 49
FT CHAIN 50 ? ADHESIN AIDA-I.
FT PROPEP 1286
SQ SEQUENCE 1286 AA; 132271 MW; B2A00F72AC05FB34 CRC64;

Query Match 4.68; Score 220; DB 1; Length 1286;
Best Local Similarity 20.0%; Pred. No. 5.3e-05;
Matches 217; Conservative 156; Mismatches 406; Indels 306; Gaps 49;

OY 36 NGINDSG-----FTPTKTSATYSLTGDVFEYEPG-----KGRP 69
DB 78 NAYVNSGGTIVVNGGTTTATTVVNSGQVNGVSGATITTVNSGGIQRVSSGGVASATN 137
OY 70 LSDSC-----FKQTTDNLFLNGHSLTFGFIDAGTH-----AGAASTTANKN 113
DB 138 LSGGAQNIVIMGHASNTVIFSGNOTIFSGGITDSTNISSGGQORVSSGGVASATN 197
OY 114 LTTSGSFLSPDSSPTVTTGGTILSSAGVNLNTRKLVAGNFTADGAIKASFL 173
DB 198 ---GAQNILSEEAISTHISISSGNQYISAGANATETI---VNSGGFORVNSGAVATGTVL 251
OY 174 LVTGSGDALSNSSSTKGAIAIT-----AGARIANNVTGGYRFLSNASTSGA 223
DB 252 SGGT-----QNVSSGGAISTSVNSGVQVFAQATVTDITTVNSGQNONISS---CG 300
OY 224 IDDEGTISLSNNKFLYFEGNA--AKTTGAICTKTASGSELI-----ISN---NKT 270
DB 301 IVSETTVNVAGTQNIYSGGSAISANIKSGQIVNSEGTAIMTVSDGGYQHIRNGIASGT 360
OY 271 LIFSNNVETSGAIIHKKLALSGGTFEFL-----RNVSSA-----T 309

```

```

DB 361 IVNOSGVYNNSSCG-YAESTIINSGLTRVLSDGYANGTILNNSGRNENGVSYNAMI 419
OY 310 PKGGAISIDASGELSTL-----AETGNITFVRNTLTGTS-TPTPRNAINISNG 359
DB 420 NTGNOYIYSDGEYFAIVNTSGFORINSGETAPVNSVYTRVSSAAPEAEVSGG 479
OY 360 KFTFLRAAKNHTTIFYDPI-----TSEGTSSDVLKINNSAGALNPYOGTI--- 405
DB 480 K-----QTVLMKGIWTSNFLTAVMSMFPPTASGA---NNVLSGRILNAFAGNYVCT 527
OY 406 -----LFSGETTLADELKVDN-----LKSFTQPVLSGK--- 437
DB 528 ILNDEGRQYVYSGATATS---TVGNNEGREVYLSGITDGVLSGGLQVNS--SGKASA 583
OY 438 LILQKGYLTLESTFSQAGSLG-----MDSGTT---LSTAGSITITMLGINVD 484
DB 584 TVINEG---GAQVYDGGGVGTINIKNGGTRVDSASALNIALSSGMLFTSTGATLP 639
OY 485 SLGLKQVSLTAKGASKRYVSGKLTLDIEGNYTESHMFSDQLFLAKITVDADVTN 544
DB 640 ELTTMALISVQNHNASIVYLENGSLAVTSGGTATDITVNSAGRL-----RIDDG 689
OY 545 VDLSLLIPVPAEDPNSEYFGQGMNV---NMTDTATNTEATATWTKGEVPSPERKS 600
DB 690 GTINGTTINAD-----GIYAGTINIONDNFLLNLNENDFETEL-----SGS 732
OY 601 ALVONTLMGVPFTDIRSLQOLVEIGATGMEHKOG--FWVSSMTN-----PLHKTGDE 649
DB 733 GVLKNDTGTMTTACITLQ-----AQGVNKNKGITIFDSAVNADMAVNONATINISDQA 787
OY 650 NRKGRHTSGYVIGGS---AHPKDDLFTFAFCHLEFARDKCEFIANNNS-----RTY 699
DB 788 TINSVNNNSIYVINSITINONITNDADLSFGTAKLSATVNGSLVNNKNIILNPTKESA 847
OY 700 GGTLEFHHSHLTPQNLRLGRAKFSASALEKPREPLALDVQVSSHNDNRMETHTYS 759
DB 848 GNLTVAS-NYTGPPGSVLSLGVLEGDNSL-----TDRLVVGKNT 886
OY 760 LPESGMSNSECIAG---TGLDLPVLSPNHPLEKTFIPOMKVEVY-----YVSONSF 811
DB 887 SGOSDIYVNEDESGGCTRGINIIISVEGNSDAEF-----SLKRYVAGAYDITLQNG-N 940
OY 812 ESSSDGRGFSIGRLNLNIPVGAKEFVQDIDGISTYDUL---SGFEVSDVYRN---NPQST 865
DB 941 ESGTDNKGW-----YLTSHLPTSDROYRPENGSVATMALANSLFLMDLNERKQFRAMSD 996
OY 866 ATLYMSPDSW-KIRGGLSLRQAFLLKGSNNY-----YVSNCELFGHYAMELR 912
DB 997 NTQPEASAVMKITGIGISSGK--LNDQONKTTTNOFTNOIGDIIKFAHQLDGFLLGIM 1054
OY 913 GSSRN 917
DB 1055 GGYAN 1059

RESULT 9
OMPB_RICTY STANDARD; PRT; 1645 AA.
ID OMPB_RICTY
AC P96989;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE (CONTRAPE: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).
GN SLP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.

```

```

Db 798 GEPPIGFAGVIVQKALLEICS---GATLTMDIDISFEHGHRTVEICNLGPAFVTGHN 854
QY 717 ---LRLGRKESSESALEKPPREIPLADVQSFSDNRMEHTHYSLPSEGSWSNE-CI 772
Db 855 TTGINSGTISLQNG--KDPAPSPYL-----LATNGSATNAGTI 893
QY 773 AGCIGIDLPFLVSNPHPLF-----KTFIPOMKVE---MYYSONSEFFESSSD---- 816
Db 894 TCKV-----TEQHSYVKNKYSTGTSNSFTFRNDVSSITGLVQNSNTITNTDGIIDL 945
QY 817 -GRGFSIGLLMLSIPIVCAKPYQGDIG--DSYTYDLSG--FEVSDDYVRRNDPSTATIV-MSP 872
Db 946 YGRG-SVGM---LAIADSTAEHQKTTIDSMWVDANDTAMRDIASNSAIDFGTGVGCT 1001
QY 873 DSWKIRGGLSR-----QAFLLRGSNNYVNS---NCELFGRH----- 907
Db 1002 DSYSGAKNKATINOLGCVITTYNAGMAAYGASNTVINGTINLEKNGNDDSLAANT 1061
QY 908 ---AMELGRSSRN-----YNDVGTGLRF 928
Db 1062 LVGMAVYEHGTAINDOTGVININVTGQAF 1091

RESULT 7
SLAP_CAUCR STANDARD: PRT: 1025 AA.
AC P35828; Q46015;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE S-LAYER PROTEIN (PARACRYSTALLINE SURFACE LAYER PROTEIN).
GN R5AA.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=76;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21: 646-660 AND 1021-1025.
RC STRAIN-ATCC 19089 / CB15;
RC MEDLINE=93007489; PubMed=1393820;
RA Gilchrist A., Fisher J.A., Smit J.K.;
RT "Nucleotide sequence analysis of the gene encoding the Caulobacter
RT crescentus paracrystalline surface layer protein."
RL Can. J. Microbiol. 38:193-202(1992).
RN [2]
RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.
RC STRAIN-ATCC 19089 / CB15;
RC MEDLINE=89008089; PubMed=3049545;
RA Fisher J.A., Smit J.K., Agabian N.;
RT "Transcriptional analysis of the major surface array gene of
RT Caulobacter crescentus."
RL J. Bacteriol. 170:4706-4713(1988).
RN [3]
RP CHARACTERIZATION.
RC STRAIN=CB15;
RC MEDLINE=98292737; PubMed=9620954;
RA Awram P., Smit J.K.;
RT "The Caulobacter crescentus paracrystalline S-layer protein is
RT secreted by an ABC transporter (type I) secretion apparatus."
RL J. Bacteriol. 180:3062-3069(1998).
RN [4]
RP FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.
CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE
CC SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A
CC SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF062345; AAC38665.1; -.
DR HSSP: P01549; 2MCM.
DR InterPro: IPR001343; -.
DR Pfam: PF00353; hemolysinCbind; 1.
DR PRINTS: PR00313; CMBDNGRP.
KW Cell wall; S-layer; Calcium-binding.
FT INIT_MET 0
SQ SEQUENCE 1025 AA; 98209 MW; AEC8B519820B1A5F CRC64;

Query Match 4.8%; Score 227.5; DB 1: Length 1025;
Best Local Similarity 20.8%; Pred. NO. 1.5e-05;
Matches 169; Conservative 107; Mismatches 273; Indels 263; Gaps 36;

QY 10 VSSVLAFCSHLSLAN---EELLSPDSEFNGINDSGTFPKTSATYSLTGDVFFYEPPGK 66
Db 205 VSGIGCYATATAAMINDSLDGLSTDNAGVNLFTAPYSSGVSGTSLT-----T 255
QY 67 GTPLSDSCFKQTTDLNLFELGN-----GHSITFEFIDAGTH---AGAAASTTANK 112
Db 256 GT-----DTLTGTANNDFEYAGVAGATLTVDGLTSGG---AGTDVLNWNVQAATAALPT 308
QY 113 NITPSCFSLSDSPSTVTGCG-----TLSSAGCVNL----- 147
Db 309 GVTISCIETMNTVSGAATLNTSSGVTGLTALNTNTSGAAGVTVTAGAGNLAATTAQAQA 368
QY 148 ENI-----RLVVAAGNFTPDGGAIGKASFLTLGTSODALFNNSSTKGAITATAGARI 203
Db 369 NNVAADGRANVTVAASGVTSCTTYTGANSAAGTYSVS--ANSTTTGALATVGTAV 426
QY 204 ANNTGYRFLSLNASTSGAIDDEGTSILSNKFLYEGNAKT-----TGCAIC 253
Db 427 -----TVAGTAGAANV---TTLTQAD--VTVTGNSSTAVTVQTAAATVATV 470
QY 254 NTKASGPELILISNNKTLIFASNVAETSGALHAKL-----ALSSGCTEFLRNV 305
Db 471 AGRVNCATVITDSAAASATTAACKIATVILGSGAATIDSSALTYVNSTGVSIGRGA 530
QY 306 SSATPKGASISIDASG-----ELISAETGNTF----- 334
Db 531 LKATPTANLTLANVGLTGTGAITDSEAAADGFTTINAGSTASTIASLVAADATTLN 590
QY 335 ---VRNTLTT-----TGSDTPPKRNA-----INISNGKTEFLRAKHHTTF 374
Db 591 ISGDARVTTTSHTAALNGITVNTNSGATLGAELATGLVFTGGAGRDSILLGATTKAIYM 650
QY 375 YDPTISEGTSDDVLAINN---GSAGALNPYOGTILFSGETLTADELKVNADNLKSSFTQPV 431
Db 651 -----GAGDPTVTVSSATLGAAGSVNGDGT-----DVLVANVGSFSADP 692
QY 432 SLSGKLLLOKGVLTLESTFESQAGSLGMDSGTTLSTAGSITTYLNLGINDSLGKOP 491
Db 693 AFGGFETTLVAGAAAG---SHNANGFTALQ---LGATAGATTTNVANVGLVTLAAP 745
QY 492 VSLT-----ANG-----ASNKVIYSGKLNIDIEGNTIESHMFSDQLFSLKIT 536
Db 746 TGTITVTLANATGSDVFNLTLSSSAALAGVALAGVE-----T 785
QY 537 VD-ADVDTN---VDISLIPVPAED--PNSYTGOGGQNNVMTDTATNTKEATATYTK 589
Db 786 VVIAATDFTTAVDTLTLQATSAKSIYVTGNAGL---NLNTNGTAVTSPDASAV--- 838
QY 590 TGFVSPERKSAVLONTLMGVFTDRLSLOLVEICATGMEHQGFVWSMTNLFKRTQDE 649
Db 839 TGTAPAVTFVSA---NTVGEVVTIR-----GAGAGDSLTG----- 871
QY 650 NRKGFRTSGGVIGGS-----AHTPKDLFT 676
Db 872 -----SATANDTIIIGAGADTLVYTGTDFTT 898

```

```

Db 1052 DGIENSTSMIVTTKANNGTIVLYGNAP-----VCNIGSDPTPV----- 1090
Qy 737 PLADVOYSSHSNRMETHYTSIPSESGSMNECTAGIGLDPFLSNPHLEKTFIP 796
Db 1091 -----ASVRETSNGNAGL-----KGNIVSOVIDEFT-YNIGYVNSVILGGSTTAI 1136
Qy 797 OMKEVYVYQNSPSESSDGRGFSIGRLNLSIPVCAKVOGDIG-----DSYT 846
Db 1137 NGKIDLTNTLTTRAGGTSTGNNTSIETTLTA-----NGTIGIYIAEGAQVATT 1188
Qy 847 YDSGFVSDVYRNPNPOSTATLVMSPDWIRG-----GNLSROAFPLRGSNNVYNSNC 901
Db 1189 TGTITINQDNANANFSGTQTYTL-----IQGAREPNTLGGPNFTVTSNRPV----- 1237
Qy 902 ELFGHYAMELRGSSRY 918
Db 1238 ----NYGL-IRANMODY 1249

RESULT 6
YDPA_ECOLI
ID YDPA_ECOLI STANDARD; PRT; 2003 AA.
AC P3666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOHETICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.
GN YDPA OR B1401/81405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SOURCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-9742617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SOURCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97251357; PubMed-9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakabe S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takekoshi K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SOURCE OF 464-2003 FROM N.A.
RC STRAIN-K12;
RX MEDLINE-92190338; PubMed-1665988;
RA Moszer I., Glaser P., Danchin A.;
RT "Multiple insertion sequences near the replication terminus in
Escherichia coli K-12.";
RL Biochimie 73:1361-1374(1991).
CC -1- SIMILARITY: NO S. TYPIMIRUM ORF NEAR CYS6 (AC P25928).
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
BETWEEN AMINO ACIDS 839 AND 840.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entries requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000237; AAC74483.1; ALT_SEQ.
DR EMBL; AE000237; AAC74487.1; ALT_SEQ.
DR EMBL; D90778; BAA15009.1; ALT_SEQ.
DR EMBL; D90778; BAA18880.1; ALT_SEQ.
DR EMBL; D90779; BAA18881.1; ALT_SEQ.
DR EMBL; X62680; NOT_ANNOTATED_CDS.
DR Ecocore; Egl1307; ydpa.
KW Hypothetical protein.
FT CONFLICT 489 489 I -> V (IN REF. 2).
FT CONFLICT 495 495 I -> V (IN REF. 2).
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;

Query Match 4.8%; Score 231; DB 1; Length 2003;
Best Local Similarity 21.4%; Pred. No. 2,3e-05;
Matches 238; Conservative 133; Mismatches 379; Indels 360; Gaps 58;

Qy 31 PDD-----SRNGNI-----DSGTFPKTSATYSLTGDFVFFPEPKGT-PLSDSCFKQ 77
Db 130 PDDTPPKPEVSEFNNVDILDKTEKTLTRDSVFYTTNAD-----GTSIQDSNGRK 181
Qy 78 TTDNLTPFL--GNCHSLTFGFIDAGT-----HAGAASTANKNLTFSGFSLSDSSPS 129
Db 182 ATTNLMQIDANNTVALEGSADCATKQYNHNGELYTGDNATVANNNGKTYDGKSTG 241
Qy 130 TTVTTGGTL-----SSAGCVNLENIRKLVYAGNESTADG-GAI----- 167
Db 242 TEINGNNGKVIQDGDLDVSGGHGID-----ITGDSATVDNKGTMVTDPESKIQIDG 295
Qy 168 -----KGSFLLTSG-----DALFSNNSSTGCAIATAGARIANNNGYVFLS 214
Db 296 DKAIYNEGSEITNGGTGTQINGDDATANNNGKTYDGKST--GTEINGNNGKVIQDG 353
Qy 215 NIATSGG-----AIDDEGTSLSNNKFLYF-----EGNAAKTTGGAIC 253
Db 354 DLVSSGGGHIDITGDSATVDNKGTMVTDPESIGIVDDQAVNNEGESAITNGT-- 411
Qy 254 NTRASGPELIIISNNKTLI-----FASNVAETSGCAIHAKKLALSSGFEFLNNVSA 308
Db 412 GTQINGDDATANNNGKTYDGKSTGTEIAGNNGKVIQDGDLDVSGGHGIDITGDSATV 471
Qy 309 TRPG-----GASIDASELSISAFGNITFYRNMLTTGSGTDTPKRAINNGSKXF 361
Db 472 DNKGTMVTDPESIGIQIDGDQAVNNEG-----ESTITNGGTGTQINGDDATANNNGK 526
Qy 362 TELRAAKNHTIFPYDPTTSGTS-----SDVLRKINNGSAGALN-PYOGTILFSGE 410
Db 527 T-----VDGKDSGTGKTAGNIGIVNLDGSLTVGGAHGVENIGDNGTVNNKGD 574
Qy 411 TLTADELKVAIDNKSSFTQPVSLSGKLLIQKGVLEST--SPSOBA-GSLGMDSGTT 466
Db 575 IIVSDTGSII-----GVLINGEGATVSNITGDVNVSNATGTFSTTNSGKV 618
Qy 467 LSTAGSITITNIGINDSLGKQPVSLTAKGANSKYI-----VSGKLNLIDIEGNI 518
Db 619 --SLAGMOVGDPSTGVDLGNNNNSVTLAKDL--KVGGKATGINSGGANVANTIGNV 674
Qy 519 YESHMSHDQLFSLKITYDADVDYTDVNDISLIPVPAE--DPNSEYGFQGGNNVNTTD 575
Db 675 L-----VDKDKTADNA-----AEYFDPDSVGIVNGSDSN-NTVLD 708
Qy 576 ---TATNTKEATVTKTGVFVSPERKSAIYC---NTLNGVFTDIRSLQQLVEIGATGM 628
Db 709 GKLTIVVSDSEVTSRQSL-EDGSAEKTSGLVIGDGTV----- 746
Qy 629 EHKQGVVSSMTNPLKRTGDN-----RKGRHHSRGYVIC-GSAHPRKDDLT 676
Db 747 -----NMNGGLELIGKKNALADGSOVTSLRGTGYSYSLVIVSGESSVYINGDTTIS 797
Qy 677 FARCHLFA--RDKCFIAHNSRTYGGTL-----FFKH--SHTLQPOVY----- 716

```

```

Db 180 NDVLSVDFDNNQAVTSTSYSDGGAIDVTDNNSDKHPGTYIVNNTA---FTNNNTA 236
Qy 246 KTTGALICNTRKASGSPDLTIS-----NKKTLIFSNVAETSGGAIHAKK 289
Db 237 EGYGGAI-VTNSVTPAPYIDISVDYSQNGVLVDENNSAAGYDGPSSAAGFMV--- 292
Qy 290 LALS-----SGTFEPFLNNVNSATPKG---GAISIDASGELSLSA-----E 328
Db 293 LGLSEVTDIADGKTLVGNTEPDGAVDSIAGTGLITKSGSDVILMNDNNFTGEMOLE 352
Qy 329 TGNITFVR-NLTITGTS---DTPKR-NAINIGSNGK---TELRAKNHTITFYDPTIS 380
Db 353 NGEVTLGRSNLSIMNVGDTHCDDPDCCYGLTIGSIDQYONQAEIWNVGSITQTFVH---AL 409
Qy 381 EGTSSDVLKINNGSAGALNPYOGCTLFEGEPTLFADELKAVADLKSSP-----TQPVSLSG- 435
Db 410 TGFQNGTINIDGGVNTVA--OGS--FAGILEGAGOLTTIAQN--GSYVLAGAQSMAITGD 463
Qy 436 -----GKLLQKGVTLTESTFSQEAAGSL-----GMDSGTTLIS 468
Db 464 IYVDGAVALSLEGDAADLTALQDDPQSYLVNGVYLDLDSFTWQSGTSYNDGLEVSQSSG 523
Qy 469 TTAGSITTTNL-----GIVNDSLGLKQPVSL-----TAKGASKNKVIYS 506
Db 524 TVIGSGQVDVADLAGDNLHIGDGKDGKGVYVVDASDQVSIANNNSYLTGTTQIASGTLAWS 583
Qy 507 GKLNIDIEGNYESHMSHQDLFSLKLTVDADVDTVDNISL---IPVPAE-DPNSKY 562
Db 584 DNSQIGDHYN--ROVITFDKQOESVMEIT--SDVTRSDAAGHRDIEMRADGEVAADA 639
Qy 563 GFGOGANNMTTDTATNTKEATATWTKTG---FVPSPERKSALVC--NTLMGVFTDIR 615
Db 640 GVDYTWGMA-LMADSSGQODESGTILTKGAGLELTAAGTSSAVRAVEGTIKGVDADI- 697
Qy 616 SLQQLVEIGATGMEHKGQFVWSMTNFLHKTG-DENRKGFRHTSGGYV 662
Db 698 -----LPPASSLWVGDAFTV--TGADQDQISIDAISSGTI 731

RESULT 5
OMP_RICJA STANDARD; PRT; 1656 AA.
ID OMP_RICJA
AC 006653;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH.
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein romp B of Rickettsia
RT japonica."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This swiss-prot entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.slb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
CC EMBL; AB003681; BAA20138.1; -.
DR Antigen; s-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
KW CHAIN 1339 1656 32 KDA BETA PEPTIDE.
FT CHAIN 528 533 POLY-GLY.
SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9D5999F CRC64;

Query Match 4.9%; Score 232; DB 1; Length 1656;
Best Local Similarity 22.2%; Pred. No. 1.6e-05;
Matches 230; Conservative 147; Mismatches 356; Indels 304; Gaps 58;

Qy 36 NGNID-----SGTFPKTSATTYSITGDFEYEDGKTPLSDFCKQT--DNLTF 84
Db 363 NGQVNFPHIYDVGIDGTTAFKTAASIVALTQNSN-----GTTDGNLAAQVTPDPTML 417
Qy 85 LGNGSLTFGFIIDAGTHAGAAASTANKNLTFSGFSLSFSDSPSTYTTGGCTLSSAG 144
Db 418 TGN-----FTGDANNPENTAGV-----ITPANGCTLASASADANVAVTNNITAEASG 465
Qy 145 VLENIR-----KLVAAGN-FSTADGAIK---ASTLITGSGDALFSNNSST--- 190
Db 466 VGVVOLSGTHIELNLGNASVFKLADGVINGKNQVVLGVGLAAGATLLDGSATYTG 525
Qy 191 ---KGAITATTAGARIANTGTVYRFL--SNIASTSGAIDDEGTSILSNKFLYEGNAA 245
Db 526 DIGNGGGGAALDITLADATITLLGCAINIISANGTIN-----EQAN-- 569
Qy 246 KTTGALICNTRKASG---PELLISNKT-LIFASNV--AET--SG--GAIHAKKALS 293
Db 570 ---GGTIKLTSTQNNIYVDCDLAIATDQGVYDASSLINAQTLTISGTIGLIANNITL- 625
Qy 294 SGGTFEFLRNNVSSA--TPKGAISIDASGELSLEAENGNTIFVNTLTGCTSTTPKRN 351
Db 626 -GQF-----NIGSSKTLNGGVAIN--ELVI--GNNSGVFAINTLYLTFTT----- 668
Qy 352 AINISNGKFTFLRAAKNHTIFFYDPTSEGTSSDVL-KINNGSAGALNPYOGTILFSGE 410
Db 669 ---NAAGQKIIFNPVYNNNTILAG--TNLSAANPLAEINSGKA---RADYLVANNGE 721
Qy 411 --TLFADELKAVADNLKSSFTQPVSLSGKLLQKGVTLTESTFSQEAG--SILGMDSGTT 466
Db 722 GVNLYATNITTTDANVGSF--VFNAAGKNI-----VSGTVGGQGGKKNFTVALDNGTT 772
Qy 467 L-----STTAGSITTTNLG-----INVDSLGLQ-----PVSLTAK 497
Db 773 VKELCNATFNGMTTLAANSTQIISGNYTADFTASDGIYEFVATGPIVTLNKAQAVP 832
Qy 498 GASNRVIVSGKLNLI-----DIEGNIYESHMSHQDLFSL-----K 534
Db 833 NALKQITVSGPQVYVNNELGNAGNHAAMTDIAEENSISLAVLFLPSGIPENDAGNTIP 892
Qy 535 ITVDADVDT-----NVDISLIPVPAE-----DPNSEYGF-OGQWN---VNWTT-- 574
Db 893 LTRKSTVENETAEGRSVSVIAGVDSYIADQVIGDQNNIYGLIGLSDNGITIVNATLY 952
Qy 575 ---DIAITNKEATATWTKTGPVPSPERKSALVCNTLMKVFDIRSLQQLVEIGAGMEHK 631
Db 953 AGIGTINNQ--GTVTLSGVPMNP-----GTVYSLGTG-----IGAS--RFX 991
Qy 632 OGFVWSSMTNF--LHKTDENKRGFRHTSGGVIGGASHTPKDGLFTFAFCH----- 681
Db 992 OVTFTTVDNNLGNITATNTTINDGYTVTTGGIAGAGIAGTDPDGATITLGSVGNANVFA 1051
Qy 682 --LFAKRDCEFL--AHNNSRTYGGTLFFKHSHTLQPNQYLRAGKAFSESAIEKFPREI 736

```

```

OY 582 -----EATATWTKGFEVSPERKALVCNTLMGVFTDRL 617
DB 1495 VLDITGDELCATTANVLENDVAQLTGTGIGFDFPNAKNGVYTLNNVNAVAGAVONT 1554
OY 618 -----QQLVEIGATGMEHKOGFVW-----SS 638
DB 1555 GGTNGTGLIVLIGASNLNRVANGIAMLKVAGCNVTIAKGKVKIGIEGTGTNTLLTPAHFN 1614
OY 639 MTNPLHTSGENKRGFRHTGGYVIG--GSAHHPKDLFPFACHLFAARKDCEFIAMNS 696
DB 1615 LTGSINTGTGQALK-LNFMNGGSVGVGTAANSVGI--TTAGATSPASS---VNAKGT 1668
OY 697 RTYGCTGFPHSHPTLPQNTLRGAKFSESATLEKPREIP---TALDVOVSFHSNDNR 752
DB 1669 ATLGCTSPANTFT-----NNGAVTLAKGSTSEFANKNTATSFVANSATINFNS--- 1718
OY 753 METHYISLPESGWSNECIAGIGIDLPEVLSPNPDLFTFIPQMKVENVYVSQNSF-- 810
DB 1719 -----LAFNSNITGGGTTTLTG--AN-----QVYTGCTGFTD 1749
OY 811 -----FESSSDRGFSIGLMLSTPYGAKFVGQDIDGSDTYDLSGFVSDYRNNPOSTAT 867
DB 1750 TLTNTTFEDGANSNGNLT-----IKSGSTLDLSGV-----STLA 1784
OY 868 LV-----MSPDS 874
DB 1785 LVYATNTFDMNNISPTD 1801

RESULT 4
YFAL_ECOLI
ID YFAL_ECOLI STANDARD; PRT: 1250 AA.
AC P45508; P45507; P45506; P39441; P76468; P77487;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 131.2 KDA PROTEIN IN UBIC-NRDA INTERGENIC REGION
PRECURSOR.
GN YFAL.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
OC [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN-K12;
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K., Kasai H.,
RA Kashimoto K., Kim S., Kikuchi S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
RX MEDLINE=84272624; PubMed=6087316;
RA Carlson J., Fuchs J.A., Messing J.;
RT "Primary structure of the Escherichia coli ribonuclease diphosphate
reductase operon.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:4294-4297(1984).
RN [4]
RP SEQUENCE OF 925-1198 FROM N.A.

```

```

RC STRAIN-K12 / EMG2;
RA Estep P., O'Keefe T., Robison K., Church G.M.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1180-1250 FROM N.A.
RC STRAIN-OV6;
RX MEDLINE=88201664; PubMed=2834621;
RA Hussain K., Elliott E.J., Salmond G.P.C.;
RT "The parA-mutant of Escherichia coli also carries a gyrAam mutation.
Mol. Microbiol. 1:259-273(1987).
RN [6]
RP IDENTIFICATION.
RX MEDLINE=96032851; PubMed=7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
three gene classes.";
RL Nucleic Acids Res. 23:3554-3562(1995).
CC -1- SIMILARITY: TO E. COLI YDEK.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
MANY FRAMESHIFTS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000313; AAC75293.1; -.
DR EMBL; D90855; BAA16052.1; ALT_INIT.
DR EMBL; D90854; BAA16050.1; ALT_INIT.
DR EMBL; K02672; -. NOT_ANNOTATED_CDS.
DR EMBL; U30459; AAA74094.1; -.
DR EMBL; Y00544; -. NOT_ANNOTATED_CDS.
DR EcoGene; EGI2850; yfal.
KW Hypothetical protein; Repeat; Signal.
FT SIGNAL 1 23
FT CHAIN 24 1250
FT DOMAIN 919 948
FT FT 28 30
FT CONFLICT 28 30
FT CONFLICT 40 40
FT CONFLICT 65 66
FT CONFLICT 431 431
FT CONFLICT 433 434
FT CONFLICT 478 478
FT CONFLICT 773 773
FT CONFLICT 853 853
FT CONFLICT 923 924
FT CONFLICT 948 994
FT SEQUENCE 1250 AA; 131152 MW; 17F98C05E299FC93 CRC64;
SQ
Query Match 5.0%; Score 240; DB 1; Length 1250;
Best Local Similarity 23.8%; Pred. No. 3 7e-06;
Matches 183; Conservative 105; Mismatches 278; Indels 202; Gaps 43;
OY 46 PKTSATYSLTGVDFVEPEKGPLSDSC--FKOTDNLN--FLNGHSLTFEFLDAGT 100
DB 15 PSMAISLFSANGVAAYVDSQGVKASQASQSLSGITQDMSADGQMLVFSMDTNN 74
OY 101 HAAGAASTANKNLTFSGFSLSPDSPT-----VYTGQGLSSAGV-----NLENT 150
DB 75 SGAVALQOG-----AEFSLPEENEGMTLFAANNVT--GEYNGGALFAENSTLNL 125
OY 151 RKLVAAGNFSTADGAIKGASFLTGTSG---DALFSNN-SSSTFGAATATAGARIAN 205
DB 126 TDVIFSGNVAAGYGALYSSTGTDGAVDLRYTNMFRNNIANDGGGALYT-----IN 179
OY 206 NTGY---VRFSLNIASTS-----GGALD-----DEGTSILSNKFLYFGNAA 245

```



```

Db 564 GTGQTPSSSTSSISTPMILFNGRLISVDENYESVYDMSDLRSGKARQLILIEFTINDGOLD 623
QY 543 TAVDLS---SLIPVPAEDPNSSEYFGOGNNVMTTDE----- 576
Db 624 SNWQSSSLNTSLSP-----HYGQGLMTPMWITTTTITLNNSSAFTSATSTAEQK 677
QY 577 -----ATNTRKATAT-----WTKTGFPVSPERKSA 601
Db 678 TSETPTPSMTTASIPNIKASAGSGSASNGSEVITFTHTLVWMAPIGYIVDPRIIRD 737
QY 602 LVCNLTMGVFTDIRLQQLVEIGATGMEHKQGFV---SSMTNPLHKTDENRKGFRHTS 658
Db 728 LIAVSL-----VHSGRRMT-MGLRSLLPDMSFALQGAATLFTKQOKRLSYHGYSSAS 790
QY 659 GGYYVGSAAHFKPKDLFFAFCHLFARPKDCFIANN---SPTYGTTLPFKSHLQPN 715
Db 701 KQYTVSSQSGAHGKHLISFSQ---SSDKMKKEKTNRLSSRYLSALCFEH----- 840
QY 716 YLRIGRAFSESSEAIKPREIPLADVOVSFSDNRMTHTYSLPESSEGSMSNECIAG 775
Db 841 -----PMEDRIALIGAACNYGTHNMSPFGTKSSGKRFHTIGAS 883
QY 776 IGLDLPVLSNPHLPKFTIPQMKVEMVYVSONSF--FESSSDGSGFSIGRLINLS--- 829
Db 884 LRCEL-----RDSMPLRISIMLPFAQALFSRREPASIRSGDLARFTLEQAH 932
QY 830 ---IPVGA-K-FVQGDIGSYTYDLSGFFSYDYRRNNPOSTATLVMSPDNRKRGGLSHQ 885
Db 933 AVVSTIGIKGAVSDPTWPLSWEMELAYOPTILMKRPLINTLILIONGWSWTTNPLAKH 992
QY 886 AFLRGSNNYVNSNCELEGHYAMELRGSSRNVNDVGTKLRF 928
Db 993 SFYRGSHSLKF-SHLKLFANYQAQAEVATSTVSHYINAGALVF 1034

RESULT 3
ID 190K_RICRI STANDARD: PRT: 2249 AA.
AC P15921:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiaceae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R;
RX MEDLINE=90354033; PubMed=2117568;
RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
RT "A protective protein antigen of Rickettsia rickettsii has tandemly
RT repeated, near-identical sequences.";
RL Infect. Immun. 58:2760-2769(1990).
CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY.
CC -1- PTM: GLYCOSYLATED (POSSIBLE).
CC -1- DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF
CC -1- AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31227; AAA26380.1; -.
DR PIR: A41477; A41477.
KW Antigen; Repeat; Signal; S-layer; Glycoprotein.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 2249 190 KDA ANTIGEN.

```

```

FT DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 212 286 A (TYPE I).
FT REPEAT 287 358 B (TYPE II).
FT REPEAT 359 430 C (TYPE II).
FT REPEAT 431 505 D (TYPE I).
FT REPEAT 506 577 E (TYPE II).
FT REPEAT 578 652 F (TYPE I).
FT REPEAT 653 724 G (TYPE II).
FT REPEAT 725 799 H (TYPE I).
FT REPEAT 800 874 I (TYPE I).
FT REPEAT 875 949 J (TYPE I).
FT REPEAT 950 1021 K (TYPE II).
FT REPEAT 1022 1093 L (TYPE II).
FT REPEAT 1094 1165 M (TYPE II).
FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
SQ SEQUENCE 2249 AA; 22433 MW; A9D6646C089DF087 CRC64;

Query Match 5.2%; Score 250; DB 1; Length 2249;
Best Local Similarity 18.5%; Pred. No. 2,3e-06;
Matches 214; Conservative 153; Mismatches 362; Indels 428; Gaps 50;

QY 24 ANEELISPDSEFNIGSIDGTFPTKTSATYSLTGDV-----FYEPRKGT-PLSDCKK 76
Db 767 ANAVLTGAIDNTTGGDNVGVL--NLNGALSQVTDIGDNTSLATISVGAGTATLGAVIK 824
QY 77 QTDNLT-----FLNGHSLFFGFLDAGTHAGAASSTANKN---LTFSSG-FILLSD- 125
Db 825 ATTKTLTNAAVSLTTLTNAAVLTGAVD-----NTTGDNVGVNLNGLASQVTDGI 875
QY 126 --SSPTVTVTGGTLLSAGVNLNIRKL-----VVAGNFSYADG-- 165
Db 876 GNTNSLATISVGAGTATLGAVIKATTKRLTNAAVSLTTLTNAAVLTGALDNTTGGDNV 935
QY 166 --AIGASFLLTGTSGD-----ALFSNNSSSTKGAIAATAGARIANTGYRFLSNLAS 218
Db 936 VLNLNGALSQVTDIGDNTSLATISVGAGTATLGAVIKATTKRLTDAASAVKFTNPVV 995
QY 219 TSGAIDDEGTSILSNKFLYEGNAKRTGGAICNTKASGSPELLISNNKTLIFASNYA 278
Db 996 T-GAIDNGN--ANGIVTFETGNS--TVGCVGNTNLA-----TVNAGAGIL 1038
QY 279 ETSGAIAHAKKLALSSGGFTFELRNNSVAPKGAIAI--SIDASGELSLSAETGNTFV 335
Db 1039 QVQGGVVKANTINL-----DNASAVFTTNVVTGAIADNG-----NANNQIVFT 1085
QY 336 RNTLTFTTGSTDTPKRNAINIGSNGKFTELRAAKNHTI-----FPYDPTTSGTSSD 386
Db 1086 GNSTVTGAVGNTNALATVNVAGLLQVOGCVVAKANTINLTDNASAVFTTNPVVVTGAIDN 1145
QY 387 VLKING-----SAGALNPYOGTILFSGETLFADELKADNLKSSFT 428
Db 1146 TGNANNGIVTFTGNTSVTGIDGNTNALATVNVAGCITTLQAGSLLAANNIDGARSLEFN 1205
QY 429 QPVSLSG- - - - -KLLQKGVLTLEST- - - - -FSOAGSLLGADS 463
Db 1206 GP--LDGGKAIPIYFKGALANGNNAIIVNTKLLTASHLTIGTVAEINIGNLF----- 1259
QY 464 GTTLESTTAGSITTN-----LGINVDSLGLKPPVSLTAKGA-SNAYI 504
Db 1260 --TIDASVGDVTLINAAQINIFRARDVVLVLSNLTGVGNMILL--AADLVAPGADEGVY 1315
QY 505 VSGKLNLDIEGNIYES-----IHMSHQLPSLKITYPAD-- - - -VDPTNVI 547
Db 1316 FNGVGNGLVGSNVACTARNIGDGGKKNFTLLIYNAVYTTDDVVLBEGIQNVLIKNADF 1375
QY 548 SSL-----IPVPAE----- 556
Db 1376 TTSSTAFNNGAIDINATYTIIDANNGNLNIIPAGNIQFAHADAQLVLQNSSGNRTITLGA 1435
QY 557 -DPNSE-----YG-----FOGQNNVMMTDTATNTK-- 581
Db 1436 IDPDNDDEGIVILNSVTAGKRLTIAGKTFGGAHKLQTLIFGAGDCS-TAGTPTNTINI 1494

```

QY	180	DALFENN-----SSSTGGAI-ATTAGARLANNTGYRFLPSTASTSGAIDDEGYSILS	233
Db	197	CELPEDNLCIOTNTNAGKGAIAAYTASNSPESNNCDL--FPIINACCGAGAIESPICSLTG	254
QY	234	NNKFLYEEGN-----AAKTGGAICNTKASGSPELLISNNKTLIPASVAETSGG	283
Db	255	NRGNIVFENNNCFKNVETASSEASDGAIKVYTRFLD---VTGNGRGRIFPESDNTIKYGG	310
QY	284	AIHAKKLLASGGFTEFLRNNVSSATPKGAISTDAGELSLAEIGNITFYVNTL---	340
Db	311	AIYAPVATLVNDGPFYFL-NNI--ANNKGAIXIYDGSNSKISADRAHIIENENITVNT	367
QY	341	-----TTGSTDPRKNAINIGNSGKFTELRAAKNTHITFYDPITSEGTSSDYLKINNGSAG	396
Db	368	NANSTISANPRRNATIVVASSSEILLGAGSSGNLFTYDIEVS-----MGVSV	418
QY	397	ALN---PYQGTILFSGETLTADDELKVAADNKKSTFOVSLSGKLLLOKGVLTLESTFSQ	453
Db	419	SPFNKADDTGVSVSGATVNSADPH-ORNIQTPTAPLTLSNGLFCIEDHAQLVNRFQ	477
QY	454	EAGSLIGMDSTTL-----STAGSTITNLTGINSL---GLKQV-----	492
Db	478	-TGGVVSIGMCAVLSCYKNGTGDASASASTLKHIGLNTSLKSGAEIPLLWEPNNNS	536
QY	493	-SLTAGSKNKFVIYSGKLNLIDIEGNT-YES---HMFSDOLEFLSKITVADVDVTND	546
Db	537	NNYATDPAATFSLSDVKLTSLIDVYDNGSPYESTDLTHALSSQPHLSTISEASDNOLOSENID	596
QY	547	ISSLLIPVPAEDPNSFYEGOGQVANNW-----TTDTATNTRKAT-----ATWTKI	590
Db	597	FSGL-NVP-----HYWMOGLMTWGMAKTODPEBASASTTIDPOKANRFHRTLLTLWPA	649
QY	591	GFVPSPEPKSALVYCNLTWLG-VFTDIRSLQDLVEIGATGMEHKOGFW---VSSMTNFIHKT	646
Db	650	GYPSPKHRSPPLIANTLTMGNMLLATELSKNSAELTPEG---HPFWCITGGGLMMVYQD	705
QY	647	GDENRKGRFRTSGGT---VIGGSANHPKCDLFTFAPCHLFAKDRCDFIAHNSRTYGGTL	703
Db	706	PREHNPGEFHMRSSGYSGAMTAGQCHT-----FSLKFSQYTKTINERY-AKNN-----	751
QY	704	FFKHSHTLOPNTYRLRLAKPSESALKEKPREIRPLADVOVSSHNSNRMETHTYSLPE-	762
Db	752	-----VSSKNYSCQCEMLFSLQ-BGF-----LTKLVGLYSTGDHNCNHHYTTGENTL	797
QY	763	-SEGSMWSECTIAGIGLDLP-VLSNHPLEKFTFIPOMKVENYVYVSONSFESSDGRGF	820
Db	798	TSGCTFRSOTWIGVAFPDLPKPRGSHIITAPRIGALG---IYSSLSHTEVGAVPRSF	854
QY	821	SIGR-LINTLSIPYAK--FVQG-DIGSXYTLDSGFVSDVYVNNPQSTATILVMSDPSK	876
Db	855	STKPLPLIVLPLVPCVKSFMNATHRPQAMTEVLA--YQPVLYROEPGIAQLLASKGIWP	912
QY	877	IRGCNLSROA	886
Db	913	GSQSPSSRHA	922
RESULT 2			
OMPF_CHLTR STANDARD; PRT; 1034 AA.			
AC	P38008;	084878;	
DT	01-OCT-1994 (Rel. 30, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	PUTATIVE OUTER MEMBRANE PROTEIN F PRECURSOR.		
GN	PMF OR CT870.		
OS	Chlamydia trachomatis.		
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
OX	NCBI_Taxid=813;		
FN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=D/uv-3/CA;		
IX	MEDLINE=99000809; PubMed=9784136;		

RA	Stephens R.S., Kalman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
RA	Mitchell W.P., Olinger L., Tatusev R.L., Zhao Q., Koonin E.V.,
RA	Davis R.W.;
RT	"Genome sequence of an obligate intracellular pathogen of humans:
RT	<i>Chlamydia trachomatis</i> .";
RL	Science 282:754-759(1998).
RL	[2]
RP	SEQUENCE OF 26-35.
RC	STRAIN=L2/434/BU;
RA	Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
RA	Comanducci M., Christlanen G., Birkelund S., Vitreou E., Ratti G.,
RA	Pallini V.;
RL	Submitted (SEP-1994) to the SWISS-PROT data bank.
CC	- I - SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES).
CC	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	
DR	EMBL; AE001360; AAC68468.1; -
FM	Outer membrane; Signal.
FT	SIGNAL
FT	1 25
FT	CHAIN 26 1034
FT	CONFICT 34 35
SO	SEQUENCE 1034 AA; 112392 MW; 445FFAC35D463AE7 CRC64;
	RR -> FH (IN REF. 2).
	PUTATIVE OUTER MEMBRANE PROTEIN F.

[illegible]

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: October 2, 2001, 03:28:05 ; Search time 48.09 Seconds

(without alignments)  
661.033 Million cell updates/sec

Title: US-09-446-677B-2  
Perfect score: 4782  
Sequence: 1 MKTSPMVLVSVLAFFSCHL.....MELRGSSRNVDVGTKLRF 928

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues  
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	610.5	12.8	964	1	OMPE_CHLTR
2	491.5	10.3	1034	1	OMPE_CHLTR
3	250	5.2	1249	1	190K_RICRI
4	240	5.0	1250	1	YFAL_ECOLI
5	232	4.9	1656	1	OMPB_RICJA
6	231	4.8	2003	1	YDPA_ECOLI
7	227.5	4.8	1025	1	SLAP_CAOCR
8	220	4.6	1286	1	AIDA_ECOLI
9	211	4.4	1645	1	OMPB_RICRY
10	209	4.4	1569	1	YFJA_ECOLI
11	208	4.3	1643	1	OMPB_RICPR
12	203	4.2	959	1	N100_YEAST
13	202	4.2	1306	1	MSB2_YEAST
14	198.5	4.2	1654	1	OMPB_RICRI
15	196	4.1	1300	1	120K_RICRI
16	191.5	4.0	1041	1	EGT2_YEAST
17	190	4.0	1150	1	APMU_PIG
18	187	3.9	881	1	YTH8_YEAST
19	185.5	3.9	2329	1	YS89_GAEEL
20	184	3.8	1039	1	AG43_ECOLI
21	183.5	3.8	1577	1	HLVA_PROMI
22	181.5	3.8	1325	1	YDEK_ECOLI
23	179.5	3.8	1113	1	N116_YEAST
24	178	3.7	995	1	Y109_YEAST
25	178	3.7	1419	1	ALAI_CANAL
26	175.5	3.7	863	1	YEJO_ECOLI
27	174	3.6	749	1	TROP_HUMAN
28	171.5	3.6	1076	1	NUPI_YEAST
29	171.5	3.6	1608	1	HLVA_SERNA
30	170	3.6	933	1	SLAP_CAME
31	169.5	3.5	1140	1	TM96_YEAST
32	169.5	3.5	1161	1	YJ9P_YEAST
33	166	3.5	2124	1	PGCA_RAT

34	163.5	3.4	1848	1	CBPA_CLOCL	P38058 clostridium
35	163	3.4	1317	1	N145_YEAST	P49687 saccharomyc
36	161.5	3.4	666	1	ALYS_ENTHR	P39046 enterococcu
37	159.5	3.3	937	1	N098_RAT	P49793 rattus norv
38	159.5	3.3	1322	1	YAG3_YEAST	P38712 saccharomyc
39	158.5	3.3	2090	1	N214_HUMAN	P33658 homo sapien
40	157	3.3	1243	1	VG37_BPK3	Q38394 bacterioph
41	157	3.3	2132	1	PGCA_MOUSE	Q61282 mus musculu
42	156.5	3.3	1849	1	IG44_HAELN	P45386 haemophilus
43	156	3.3	3591	1	FHAB_BORPE	P12255 bordetella
44	155.5	3.3	917	1	HKA3_HAELN	P45355 haemophilus
45	154.5	3.2	671	1	ALYS_ENTRA	P37710 enterococcu

## ALIGNMENTS

RESULT	ID	OMPE_CHLTR	STANDARD	PRT	964 AA.
AC	084877				
DT	30-MAY-2000	(Rel. 39, Created)			
DT	30-MAY-2000	(Rel. 39, Last sequence update)			
DT	30-MAY-2000	(Rel. 39, Last annotation update)			
DE	PUTATIVE OUTER MEMBRANE PROTEIN E PRECURSOR.				
GN	PMPE OR CT869.				
OS	Chlamydia trachomatis.				
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.				
OX	NCBI_TaxID=813;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-D/UW-3/CX;				
RX	MEDLINE=99000809; PubMed=9784136;				
RA	Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,				
RA	Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,				
RA	Davis R.W.;				
RT	"Genome sequence of an obligate intracellular pathogen of humans:				
RT	Chlamydia trachomatis."				
RL	Science 282:754-759(1998).				
CC	-I- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)				
CC	(POTENTIAL).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	at the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL: AE001360; AAC68467.1; -				
KM	Outer membrane; Signal.				
FT	SIGNAL				
FT	CHAIN				
FT	19 964 PUTATIVE OUTER MEMBRANE PROTEIN E.				
FT	SEQUENCE				
FT	964 AA; 104703 MW; 1B998A7D2E571CE2 CRC64;				

Query Match 12.8% Score 610.5; DB:1; Length 964;  
Best Local Similarity 26.1% Pred. No. 2.5e-27;  
Matches 253; Conservative 146; Mismatches 360; Indels 211; Gaps 46;

QY	45	TPKATATYSLGDFEYEPKGPGLSDSCFQKOTDNTFLGNSLTFGFIADCTHAGA	104
DB	36	TKESLSNKRISLTGD-----TINLTNCR-----LDNLRIT-----LAIQKTPNPGA	76
QY	105	AASTT-----ANKNLFGSGFSLISFDSSTTV-----TTG-----	135
DB	77	AVTIDIVLSEFDQKEGYFAKNLTPESGGAIGYASPNSPVEIRDTIGPIFENNCCR	136
QY	136	-----QGLSSAGGVNLENT---RKLTVAG---NFSTADGAATGA-SFLTGTSG	179
DB	137	LEFWNPPYAADKIRGGAIAHONLXINHNDVYGMKNESYVGGAISTANTFVASENOS	196

Db 358 KNLISGTTIILDPGTFEYENHSLRNPDSYDILELKAAGTSTS---TAVTPDPIMGKEK 413  
 QY 561 EYGFQGMGN-VNMTTDPATNTKEATATWTCTGVPSPERKSALVNTLMGVEFDIRSLQ 619  
 Db 414 HYGQGTMGPTLVKGTGAST---ATFNWTKGTIIPERIGSLVPSNLMAFIDISSLAY 470  
 QY 620 LVEIGATGMEHKGQFVWSSMTNPLFKTGDENRKGRFHTSGYVIGSAAHPPKDLTFPAF 679  
 Db 471 LMETANEGLOQDRAFCAGLSNPFHKNOSTKTRRGFRHLSSGYVIGNLHTCSDLILSAF 530  
 QY 680 CHLHARPKDCIANNRSRTYGTGTFEKKSHTLQPNQTLRLGRAFSESALKEPREIPLA 739  
 Db 531 COLHGRBDYVAAKNOGTGGYGLTYQHNET---YISL-PCKLRCPSLSYVTEIPVL 584  
 QY 740 LDVQVFSHSDNRHETHTSLPSESGSMNCEIAGIGLDLPFVLSNPHLPKFEIPOMK 799  
 Db 585 FSGNLSTHTNDKLTATYTYPTKSGMNDSPFLERCGRAPITL-DESALFEDYMEPMK 643  
 QY 800 VEMVYVSONSEFFESSDGRGFSIGRLNLSTIPGAKF-VQGDIDSYTYDLSGFFVSDVY 858  
 Db 644 LQFYVAHQEGFEKGTAREGSSRLVNLALPIGIRFDKESDQDA-TYNLTIGTYDVLV 702  
 QY 859 RNNQSNATLWSPDKIRKINGNISROAFLLRGSNNTVYNSNCELFHYAMELRGSSRN 918  
 Db 703 RSNPDCCTTTRISGDSWKTGCTNLARQALVLRAGNHCFNSFEAFSQFSELRGSSRN 762  
 QY 919 NVDVGTCLRF 928  
 Db 763 NVDGAKYQF 772

## RESULT 15

E72130  
 polyomphic membrane protein G family CP0761 [imported] - Chlamydia pneumoniae (stra  
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
 C:Accession: E72130; G81541  
 R:Kalmun, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
 Nature Genet. 21, 385-389, 1999  
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A:Reference number: A72000; M01D:99206606  
 A:Accession: E72130  
 A:Molecule type: DNA  
 A:Residues: 1-841 <ARN>  
 A:Cross-references: GB:AE001586; GB:AE001363; NID:94376263; PIDN:AA018172.1; PID:9437627  
 A:Experimental source: strain CM029  
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Ginn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.  
 A:Reference number: A81500; M01D:20150255  
 A:Accession: G81541  
 A:Molecule type: DNA  
 A:Residues: 1-841 <REA>  
 A:Cross-references: GB:AE002235; GB:AE002161; NID:97189672; PIDN:AAF38561.1; PID:9718967  
 A:Experimental source: strain AR39, HL cells  
 C:Genetics:  
 A:Gene: pmp\_2; CP0761

Query Match 30.0%; Score 1435.5; DB 2; Length 841;  
 Best Local Similarity 35.9%; Pred. No. 1.5e-73;  
 Matches 340; Conservative 147; Mismatches 335; Indels 125; Gaps 15;

QY 1 MKTSPWLVSVIATF-CHLOSIANEELLSPDPSFNGNIDSGTFMPKPTSA----TYSYL 55  
 Db 1 MKTLPRLILSLVPTLSMNLGAATTEBELSASNSFDGTTSTTSFSSKTISSATMDGTYVAF 60  
 QY 56 TGVYFFE-PGKGTPLSDSCFKQ--TTDNLTFGLNGHSLTFGFIDAGTHAGAA-ASTTAN 111  
 Db 61 KDSVVIENPKTGTGTSCFKNDAAAGDLNLFJGGFSFTFSNIDATVTASGAAGSEAN 120  
 QY 112 KNTLFGGSLSDSSPSTVTYTGCTLSSAGVNLNTRKLVVAGNFSTADGAIKAS 171

Db 121 KTVTLGSFALSFLKSPASTYTNGLGAINVKGNLSLDNDKVLQDNFSTGDDGAINCAG 180  
 QY 172 FLTGTSGDALLFSNNSSSTKGAIAATAGARIANNNGTVRFSLNIASSTGCAIDDEGTST 231  
 Db 181 ----- 180  
 QY 232 LSNKKFLYFEGNAKATTGGAICNTRKASGPPELLISNNKTLIFASNVAFETSGAIHAKKA 291  
 Db 181 -----SLKIANKSLSFIGNSSSTKGAIAHTNLT 210  
 QY 292 LSSGFEFLRNNVSSATPPGCAISIDASGELSASFTGNTTFVKNLTFTGSTDTPKRN 351  
 Db 211 LSSGFEFLFOGNAPTAAGGAGAIADSGTLSTSGSDIIEFGNTL---GAGCTYSHS 267  
 QY 352 AINISGKTELRANNNHIFPDPTTSEGTSS--DVLKINNSAGALNDYOGTILFSG 409  
 Db 268 AIDGTSAKITLALMAAGHTIYFPDPTTVTGSTSVADALNINSPDQDNKEXTIYFSG 327  
 QY 410 EFTLADLKVADNLKSFQTPVSLSGKLLQKGVLTLESTFSQEGASLIGMDSGTLST 469  
 Db 328 EKLTFAKDEKNTSKLIDNVAKKNTVVLKGDVLSANGFSDANSKLIIMDGTSLVA 387  
 QY 470 TAGSTITNLGINVDSLGLKQPVSLFAKASNKVIYSGKLNLDIEGNIYESHMFSDOL 529  
 Db 388 NTFSEIELFNLNIDSLRNGKKIKLSAATQKDIRIDRPVVALISDESFYQNGFLNEDHS 447  
 QY 530 F-SLKTTVDADVNDVNDISLIPVPAEDPNSEYFQCGQNVNMTDTATNTKEATATWT 588  
 Db 448 YDILELDAKDKDIYISADSRIDAV-----QSPYQYQKWTINWSTD---DKKATVMA 498  
 QY 589 KTGVPSPERKSALVCTLWGVFTDIRSLQIQLVEIGATGMEHKGQFVWSSMTNPLHKTGD 648  
 Db 499 KOSTNPFAOEAPLPVPLWLSFTIVASFOFTLGTGAPYKREFFVAVISVNLHRSR 558  
 QY 649 ENRKGFRHNSGGVYVIGSASATPKDILFTFAFCHLFARDKCFIAHNSRYGTLTFKHS 708  
 Db 559 ENGRKFRHVSAGVAVGASTMPGGDITLSLGFADLFANDKQFYNNTRNAKTYASLSLRQHD 618  
 QY 709 HTLOPQNYLRLGRKFSESAIEKFPREIPLALDVOVSFSDNRHETHTSLP----- 761  
 Db 619 ASLYSVVSTILGCGLEILLPVYSKTLPCSFYGOISYGHTDHRMKE--SLPPPPPLTS 676  
 QY 762 ESEGSMSNECIAGIGLDLPFVLSNPHLPKFTIPQKVMYVYVSONSEFFESSDGRFS 821  
 Db 677 TDHTSMGGYVWAGELGTRAVENTSGRFGFQETTPYKQAVYARQDSFVELCAISRFS 736  
 QY 822 IGRLLNLSIPVGAKEVQGDIGDSTYTYDLSGFVSDVYRNNPOSTATLVSPDSMKIRGN 881  
 Db 737 DSHLYNLATPLGIR-LEKRFABQY-YHVYVAMYSPDVCRSNPKCTTLLSNGSMKTKGSN 794  
 QY 882 LSRQAFLLRGSNNVYVNSNCELFHYAMELRGSSRNVDVGTCLRF 928  
 Db 795 LARQAGIVQASGRFSIGAAALFNGFGEWGRSSRSINVDAGSKIRF 841

Search completed: October 2, 2001, 03:28:38  
 Job time: 1295 sec

QY 881 NLSRQAFLLRGSNYYVNSKCELFHYAMELRGSSRNIVDYGTKLRF 928  
||:| | : : : : | : : | : | : : : : |  
Db 889 NLARQGSVRANHFQVNPHELFQFAFEVRSSSRNNTNLSKFCF 936

### RESULT 13

polymorphic membrane protein G family CP0308 [imported] - Chlamydophila pneumoniae (strain AR39)  
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: B61591  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heldelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255  
A:Accession: B61591  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-936 <REF>  
A:Cross-references: GB:AE002193; GB:AE002161; NID:97189234; PIDN:AAF38165.1; PID:97189234  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
/:Gene: CP0308

Query Match	36.28;	Score 1733;	DB 2;	Length 936;
Best Local Similarity	39.98;	Pred. No. 2.6e-90;		
Matches 378; Conservative	182;	Mismatches 356;	Indels 32;	Gaps 18

1 MKTSIPWLVSSVLAFCSCHLOSIA NEELISPDOSFNGNIDSGTF-----TPKTSATTYSL 55

Db 1 MKSSVSNIFFSSSIPIFFSSLSIVAAEVTLDSSNNSYDGS - NGTTFVFTTDAAGCTTSL 59

QY 56 TGDVFEYEPGK-GTPDSDSCFQOTDTNLELGNHSLTFEGELIDAGTHAGAASTT-ANKN 11  
||| | ||| : : ||| | : ||| : ||| : ||| : ||| : ||| : |||

Db 60 LSDVSFQNALGIPLASGCFLEAGGDLTFQGNQHALKFAFINAGSSAGTVASTADKN 119

114 LTFSGFSLFSDSSPTVIF-IGGGITSSAGGVNLENIRKLVAAGNFSDMGCAIKGASE 174

173 LITGTSGDALFSSNNSSST--KGAIATTAGARIANTGYVRETSNIASSTGGAIDDEGTS 233

Db  
180 LLSGTSGFASFSRNDAAFTGKGCGVYAATGTTITENSPGIVSFSNLAKSGSGLYSTDNC 230

QY 231 ILSNNKFLYFEGN---AAKTGGAICNFKASGSPELLISNNKTLIFASNVAETSGAII 286

Db 240 SITDNFOYIFDGSNSAMEAAQAQGAICCTTTDKT--VTLTGKNKLSFTNNIATLTGGSAIS 29

287 AKKLATSSGGFTFLRNWSSATP--KGAISIDAGSELSTAGTGNLFVVRNTLLPTTG 34

Db 298 GLKVSISAGPTLF-QSNISGSAGQGGGCAINIASAGELALSTSGDITF-NNNQYING 35

[illegible]

05 702 TLEFKHSHTL-OPONYLEBGRAKFSESATIEKEPREIPLADVOSFSHSDNRMETHYTSL 760

Db 711 SLVFHHTEGLEFDINFL-WGATRAPVULSEISQIIPLSFDAKFSYLTDNHMKKYYTDN 769

QY 761 PESEGSNNECIAGIGLDLPVLSNPPLFKTFIPQMKEWVYVSQNSFFESSDGRGE 820

Db 770 SIKGSWRNDAFCADLGASLPVISVPY-LKKEVEPFVKVQYIIYAHQDDFYERYTAEGRAF 828

829  
nh

Qy	88.1	NLSROAFLLRGSSNNYVYNSNCELEGFHAMELRGSSRNYNVDVGTGLRF	928
Db	88.9	NLARQGFVFRANHFQVNPHEIFQGFEEFVRRSSRNNTNLGSKFCE	936
RESULT	14		

C/date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2000  
C/accession: H86492  
R/Shtalr, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.;  
Suzuki, T.; Nakamura, S.; Taniguchi, Y.; Nishimura, Y.; Nakagawa, M.;

A; Title: Comparison of whole genome sequences of chlamydia pneumoniae J138  
Nucleic Acids Res. 28, 2311-2314, 2000  
A; Reference number: A86491; MIMD-20330349

A/Accession: H86492  
A/Status: preliminary

A;Molecule type: DNA  
A;Residues: 1-772 <STO>

A; Cross-references: GB:BA000008; NID:g89/8389; PIDN:BAA98226.1; GSPDB:GNCU  
A; Experimental source: strain J138

A;Gene: pmp\_3\_2

Query Match 33.3%; Score 1591; DB 2; Length 772;

Best Local Similarity 42.7%; Pred. No. 2.1e-82;  
Matches 337; Conservative 146; Mismatches 275; Indels 32; Gaps

Db 1 MKSSVSWLFFSSILPFLSSLSIVAAEVLTDSSNNSDGS-NGTTFVSTTDAAGTYYSL 59  
 QY 56 TGDVFEYEPK-GTPLSDSCFKOTDNLTFNGHSLTFPGIDAGTHAAGASTT-ANKN 113  
 Db 60 LSDVSPONAGALGIPLASGCGLEAGDGLTFQGNHAKLFAFINAGSSAGYASTADKN 119  
 QY 114 LTFSGFLSFDSSPSTTVT-TGQGLTSAGVNLNIRKLVVAGNESTADGAIKASF 172  
 Db 120 LTFNDFRLSLTISCPSLLSPGTCALKSVGNLSLTGNSQIIFQNFSSDNGVYINTKNE 179  
 QY 173 LMTGSDALFSSNNSST--KGAIAITTAGARIANNNGYRFLSNLSTSGAIDEGTS 230  
 Db 180 LLSGTSPASFASRNQAFATGKGGVYATGTTIENSPIVFSQNLAKGSGGALYSTDC 239  
 QY 231 ILSNNKFLYEFGN---AAKTGTGAICNTKASGSPELIISNNKTLIFASVAETSGAIH 266  
 Db 240 SITDNFQVIEDGNSAMWAAQAOGALICCTTDTK--VTLLGNKMLSTNNFTALYGGALS 297  
 QY 287 AKKLALSSGGFTFLRNWSSAP--KGAISIDASGELSLSAETGNITFVRNTLTG 343  
 Db 298 GLKVISAGPPLTF-QSNISGSSAGOGGAINIASAGELALSTSGDITF-NNQVYNG 355  
 QY 344 STDPKRNAINISNGKFTELRAAKNHTIFFYDPTSEGT--SSDYKINKSGAGALNPY 401  
 Db 356 STST--RNAINIITDIAKVTISRAATGOSIFYDPTTPGTAASDTLNLMLADANSEIEY 413  
 QY 402 QGTILFSGEFLTADKLVANLKSFTQPVSLSGKLLLOKGYLTLESTSPSOEAGSLGM 461  
 Db 414 GGAIVFSGEFLSPEKAIANVSTIRQPAVLARGDLVLDGYTVYFPKDLTQSGSLIM 473  
 QY 462 DSGTTLSTAGSITTNLGINVDSLAKOPVSLTAKGASKYVSGKLLIDIEGNYES 521  
 Db 474 DSGTTLTSAKANSLINGLAVALNLSLDTNKALKTEADKNISLSTGIALIDEGSFYEN 533  
 QY 522 HNSHQLESLKITVDADVTDVNDISSILPVAEDPNSRYGOGQMNVMWTTDTATNK 561  
 Db 534 HNKSASTYPLLELT-AGANGTITLGAISTLLOEETHYGGQNMQLSMA--NATSSK 590  
 QY 582 EATATWKTGFVSPERKSAVLCNTLWGYTDIRSLQQLVEIGATGMEHKGQVWSSMTN 641  
 Db 591 IGSINMTRGVIISPEKSNLPLNSLMGNFIDIRSLNQLLETSSGEPFERELMISGIAN 650  
 QY 642 FLAKTGDNKGRHRTSGYVIGSAHTPKDILPTAFCHLFAKDCFTAHNNSRTYGG 701  
 Db 651 FFRYRDSMPTRHGRHRTSGYVIGSAHTPKDILPTAFCHLFAKDCFTAHNNSRTYGG 710  
 QY 702 TLFKHSHTL-QPONTLRIGRAKFSSEAIKFRREIPLADVOVSPSHSNRRETHYSL 760  
 Db 711 SLTFHHTREGFLDIANTL-WGKATRAPVWLSEISQIILPLSDAKFSLYLTNDHMKTYTDN 769  
 QY 761 PESEGSWNECIAGIGLDLPVLSNPHLPFKFTIPQMKVEMVYVSQNSPFESSSDGRG 820  
 Db 770 SILKSGMRNDAFCADGASLPEYISVY-LLKVEEPVVKQYIYAHQDQYERKHAERAF 828  
 QY 821 SIGRLNLSTIPVAKVQGDIGSYTYDLSGFFVSDVYRNPNPOSTATLVNPSDKIRG 880  
 Db 829 NKSELLNVEIPIGVTERDSKSEKGYDILTIAMYLDAVRNPKQCSLIASDANMAAYGT 888  
 QY 881 NLSROAFLRGSNNYVYNSCELFCHYAMELKRSSRYNANDVGTKLRF 928  
 Db 889 NLRQGSVRAANHFQVNPHEIFGQFAFEVRSSRYNTNLSKCF 936  
 RESULT 12  
 C86546  
 polymorphic outer membrane protein G family [imported] - Chlamydia pneumoniae (strain  
 C:species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
 C:accession: C86546  
 C:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A:title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A:reference number: A86491; M01D:2030349

A:Accession: C86546  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1936 <STO>  
 A:Cross-references: GB:BA000008; NID:98978817; PIDN:BA98653.1; GSPDB:GN00142  
 A:Experimental source: strain J138  
 C:Genetics:  
 A:Gene: pmp-7  
 Query Match 36.2%; Score 1733; DB 2; Length 936;  
 Best Local Similarity 39.9%; Pred. No. 2, 6e-90;  
 Matches 378; Conservative 182; Mismatches 356; Indels 32; Gaps 18;  
 QY 1 MKSIPVAVVSSVLAAPSCHQSLANELLSPDDSFKNIDSGF----TPKTSATYYSL 55  
 Db 1 MKSSVSWLFFSSILPFLSSLSIVAAEVLTDSSNNSDGS-NGTTFVSTTDAAGTYYSL 59  
 QY 56 TGDVFEYEPK-GTPLSDSCFKOTDNLTFNGHSLTFPGIDAGTHAAGASTT-ANKN 113  
 Db 60 LSDVSPONAGALGIPLASGCGLEAGDGLTFQGNHAKLFAFINAGSSAGYASTADKN 119  
 QY 114 LTFSGFLSFDSSPSTTVT-TGQGLTSAGVNLNIRKLVVAGNESTADGAIKASF 172  
 Db 120 LTFNDFRLSLTISCPSLLSPGTCALKSVGNLSLTGNSQIIFQNFSSDNGVYINTKNE 179  
 QY 173 LMTGSDALFSSNNSST--KGAIAITTAGARIANNNGYRFLSNLSTSGAIDEGTS 230  
 Db 180 LLSGTSPASFASRNQAFATGKGGVYATGTTIENSPIVFSQNLAKGSGGALYSTDC 239  
 QY 231 ILSNNKFLYEFGN---AAKTGTGAICNTKASGSPELIISNNKTLIFASVAETSGAIH 266  
 Db 240 SITDNFQVIEDGNSAMWAAQAOGALICCTTDTK--VTLLGNKMLSTNNFTALYGGALS 297  
 QY 287 AKKLALSSGGFTFLRNWSSAP--KGAISIDASGELSLSAETGNITFVRNTLTG 343  
 Db 298 GLKVISAGPPLTF-QSNISGSSAGOGGAINIASAGELALSTSGDITF-NNQVYNG 355  
 QY 344 STDPKRNAINISNGKFTELRAAKNHTIFFYDPTSEGT--SSDYKINKSGAGALNPY 401  
 Db 356 STST--RNAINIITDIAKVTISRAATGOSIFYDPTTPGTAASDTLNLMLADANSEIEY 413  
 QY 402 QGTILFSGEFLTADKLVANLKSFTQPVSLSGKLLLOKGYLTLESTSPSOEAGSLGM 461  
 Db 414 GGAIVFSGEFLSPEKAIANVSTIRQPAVLARGDLVLDGYTVYFPKDLTQSGSLIM 473  
 QY 462 DSGTTLSTAGSITTNLGINVDSLAKOPVSLTAKGASKYVSGKLLIDIEGNYES 521  
 Db 474 DSGTTLTSAKANSLINGLAVALNLSLDTNKALKTEADKNISLSTGIALIDEGSFYEN 533  
 QY 522 HNSHQLESLKITVDADVTDVNDISSILPVAEDPNSRYGOGQMNVMWTTDTATNK 561  
 Db 534 HNKSASTYPLLELT-AGANGTITLGAISTLLOEETHYGGQNMQLSMA--NATSSK 590  
 QY 582 EATATWKTGFVSPERKSAVLCNTLWGYTDIRSLQQLVEIGATGMEHKGQVWSSMTN 641  
 Db 591 IGSINMTRGVIISPEKSNLPLNSLMGNFIDIRSLNQLLETSSGEPFERELMISGIAN 650  
 QY 642 FLAKTGDNKGRHRTSGYVIGSAHTPKDILPTAFCHLFAKDCFTAHNNSRTYGG 701  
 Db 651 FFRYRDSMPTRHGRHRTSGYVIGSAHTPKDILPTAFCHLFAKDCFTAHNNSRTYGG 710  
 QY 702 TLFKHSHTL-QPONTLRIGRAKFSSEAIKFRREIPLADVOVSPSHSNRRETHYSL 760  
 Db 711 SLTFHHTREGFLDIANTL-WGKATRAPVWLSEISQIILPLSDAKFSLYLTNDHMKTYTDN 769  
 QY 761 PESEGSWNECIAGIGLDLPVLSNPHLPFKFTIPQMKVEMVYVSQNSPFESSSDGRG 820  
 Db 770 SILKSGMRNDAFCADGASLPEYISVY-LLKVEEPVVKQYIYAHQDQYERKHAERAF 828  
 QY 821 SIGRLNLSTIPVAKVQGDIGSYTYDLSGFFVSDVYRNPNPOSTATLVNPSDKIRG 880  
 Db 829 NKSELLNVEIPIGVTERDSKSEKGYDILTIAMYLDAVRNPKQCSLIASDANMAAYGT 888

	Query Match	36.8%	Score 1761;	DB 2:	Length 930;	
	Best Local Similarity	41.7%	Pred. No. 6.7e-92;			
	Matches 394;	Conservative 165;	Mismatches 355;	Indels 30;	Gaps 14;	
QY	1 MKTSLPWLVVLSVLAFSCHLQSLAN---EELSPDSDSENGNIDSGTFPPKTSA----TTY	53				
Db	1 MKPIPLHLISLTIVTP-I-LTSLIATFYGADASLPSPDFDG--AGGSTFPFKSTADANGIN	58				
QY	54 SLTGCVFFEYEFGCKGPLPDSCEPKOTTDNLTEFLGNCHSLTFPGFIDAGTAGAAASTTANKN	113				
Db	59 VLSGVNVINDAGKKGALPGCCFTETTTGGDTLEPTGKYSSFFPMVDGNSAGAASAATPYAKKA	118				
QY	114 LTFEGSFLLSDSSPTVTMTGGGLSSAGGVENLNRIKLIVAGNFSA---DGAIGKA	170				
Db	119 LTFEGFSMLSTIAAGCTIVASGKSTLSSAGALNDLTONCITLFPSQVNSEANNNGCALITYK	178				
QY	171 SELLGTSGDALFSNNSSSRTKGAIATATTAGARIANTGYVRFLSNIASTSGAIDDECTS	230				
Db	179 TLTSGNGNTSITFTFSNAKKILGAIYSSAASISGNTQLVFMMNNKGFTGGALGFEEASS	238				
QY	231 TLSSNKELYEGNAAKTY--GGAICNTKASGSPELLISNNKTLLIFASNVAETSIGAIIH	287				

[illegible]



polymorphic outer membrane protein G family [imported] - Chlamydia pneumoniae (strain C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: D86546  
R:Shirai, M.; Hiraoka, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: D86546  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-930 <STO>  
A:Cross-references: GB:BA000008; NID:98978818; PIDN:BA98654.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: pmp\_8

Query Match 36.9%; Score 1763; DB 2; Length 930;  
Best Local Similarity 41.7%; Pred. No. 5.2e-92;  
Matches 394; Conservative 165; Mismatches 355; Indels 30; Gaps 14;

QY 1 MKRSIPWVAVSVLAFSCHQSLAN---EELSPDPSFNKGNIDSGFTTPTSA-----TTY 53  
DB 1 MKTLPKHLISSTLVPTI-LLSIATYGAASLSPDPSFDG-AGGSTPTPKSTADANGTNY 58  
QY 54 SLTGDFVEYEPGKGTPLSDSCFKQTTDNLFLNGHSLRFGFIDAGTHAGAATSTANKN 113  
DB 59 VLSGANYINDAGKGTALGCGFTTGTDLFTFKGYSFSFNTRYDAGSNAGAAATRADKA 118  
QY 114 LTFSGFSLSDSSPTTYYTGGTSSAGCVNLEIRKLIVAGNFSTA--DGAIRKA 170  
DB 119 LFTFGSNLSFIAFGTYVAGSKSTLSSAGALNLTDNGTLLFQSNVSEANNGAATK 178  
QY 171 SFLTGTSGDALFSSNNSSTKGAIAITTAGARTANNTGYREFLSNASTSGAIDDEGTS 230  
DB 179 TLSISGNTSITTSNASKLGAITSAAASISGNTGQVFNNGGEGGALGEASS 238  
QY 231 ILSNNKFLYEGNAATTT--GGAICNTRKASGPELISNNKTLIFASNAEFTSGAIIHA 287  
DB 239 SITQNSLTFSGNTATDAGKGAICYCEKTEGPTLISNKKSLTFAENSVYOGAIIA 298  
QY 288 KKLALSSGGTEFLRNWV--SSATPKGGAISIDASGELSIAETGNTTFRVNTLYTTGSTD 346  
DB 299 HGLDLSAAGPTLFSNNRCGTAAGKGAIALIADSGSLISANOGDITFLGNTLTST-SAP 357  
QY 347 TPKRNAINISNGKFTLEIRAKNHTTFFYDPTISEGT--SSDVKINNKGAGALNPQGTI 405  
DB 358 TSTRNATVYSSSAKITNLRRAOGOSTIFYDPLASNTTGADVLITINQPPNSPLDYSGTI 417  
QY 406 LFSGETTLADELKVADNLKSSFTQPVSLSGSKLLQKGYLLESTFSQEGASLGMDSGT 465  
DB 418 VFSGEKLSADEKAAADNFSLIKQPLASGTLAKGNVELDNGTQTEGSLTMQPGT 477  
QY 466 TLTSTAAGSTITTNLGINVDSLKQPVSLTAKGASKNVIVSGKLNLDIEGNTYESHMS 525  
DB 478 KTKADTEASLTKLVYDALBLEGNKSVSIETAGANKTITLTSPVLVQDSSGNGYFESHIT- 536  
QY 526 HDQFSL-LKITVDADVNDVNDISLIPVPAEDPNSEYGGOGOMNNTTTPATANKKAT 584  
DB 537 -NOAFQPLVFTAAATASDIYIDALLTSPVQTPPEPHYQOGHEWTW-ADIST-AKSGT 593  
QY 585 AMWTKTGFVSPERKSAVNCNTLMGVFTDIRSLQOLVEIGATGMEHKGQFWVSSMTNPLH 644  
DB 594 MMVVTGYNPNPERRASVYVDSLMASTFDIRTLQJMTSQANSIYQGRGLMAGSCTANFPH 653  
QY 645 KTGDENKRRGFRHISGCVIVGSAHPTKDLFTFAFCHLARDKCFIAHNSRTYGGTLE 704  
DB 654 KDKSGTNOAFRRHRSXYIVGGSADFSENIFSAFCOLFGKXKDLFIYNTSHNYLASLY 713  
QY 705 FKHSHTLOPNYRLIGRAFSSEAIKPPRETPALDVVSPSHNRMEHTYSLPESE 764  
DB 714 LQHRALG-----GLPMPFSGSITDMKLDLPILNLAOLSYSTYKNDMDTRTYSTPEAQ 766

polymorphic outer membrane protein G family CP0307 [imported] - Chlamydia pneumoniae (s  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: AB1591  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Grimm, M.; Nelson, W.; Debby, R.; Kolonay, J.; McClarty, G.; Salze  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
A:Reference number: AB1500; MUID:20150255  
A:Accession: AB1591  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-930 <REAS>  
A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AAF38164.1; PID:g718  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: CP0307

Query Match 36.9%; Score 1763; DB 2; Length 930;  
Best Local Similarity 41.7%; Pred. No. 5.2e-92;  
Matches 394; Conservative 165; Mismatches 355; Indels 30; Gaps 14;

QY 1 MKRSIPWVAVSVLAFSCHQSLAN---EELSPDPSFNKGNIDSGFTTPTSA-----TTY 53  
DB 1 MKTLPKHLISSTLVPTI-LLSIATYGAASLSPDPSFDG-AGGSTPTPKSTADANGTNY 58  
QY 54 SLTGDFVEYEPGKGTPLSDSCFKQTTDNLFLNGHSLRFGFIDAGTHAGAATSTANKN 113  
DB 59 VLSGANYINDAGKGTALGCGFTTGTDLFTFKGYSFSFNTRYDAGSNAGAAATRADKA 118  
QY 114 LTFSGFSLSDSSPTTYYTGGTSSAGCVNLEIRKLIVAGNFSTA--DGAIRKA 170  
DB 119 LFTFGSNLSFIAFGTYVAGSKSTLSSAGALNLTDNGTLLFQSNVSEANNGAATK 178  
QY 171 SFLTGTSGDALFSSNNSSTKGAIAITTAGARTANNTGYREFLSNASTSGAIDDEGTS 230  
DB 179 TLSISGNTSITTSNASKLGAITSAAASISGNTGQVFNNGGEGGALGEASS 238  
QY 231 ILSNNKFLYEGNAATTT--GGAICNTRKASGPELISNNKTLIFASNAEFTSGAIIHA 287  
DB 239 SITQNSLTFSGNTATDAGKGAICYCEKTEGPTLISNKKSLTFAENSVYOGAIIA 298  
QY 288 KKLALSSGGTEFLRNWV--SSATPKGGAISIDASGELSIAETGNTTFRVNTLYTTGSTD 346  
DB 299 HGLDLSAAGPTLFSNNRCGTAAGKGAIALIADSGSLISANOGDITFLGNTLTST-SAP 357  
QY 347 TPKRNAINISNGKFTLEIRAKNHTTFFYDPTISEGT--SSDVKINNKGAGALNPQGTI 405  
DB 358 TSTRNATVYSSSAKITNLRRAOGOSTIFYDPLASNTTGADVLITINQPPNSPLDYSGTI 417  
QY 406 LFSGETTLADELKVADNLKSSFTQPVSLSGSKLLQKGYLLESTFSQEGASLGMDSGT 465  
DB 418 VFSGEKLSADEKAAADNFSLIKQPLASGTLAKGNVELDNGTQTEGSLTMQPGT 477  
QY 466 TLTSTAAGSTITTNLGINVDSLKQPVSLTAKGASKNVIVSGKLNLDIEGNTYESHMS 525

```

QY 175 TGTSGDALFSSNNSSST---KGAIAATAGARIANTNTGYVRFSLNIASTSGAIDDEGTSI 231
DB 180 KNSGTSISFEGKSSAATCKKGALCATGTVDITNTTAPTLFSNNIAEAGAAGINSTGCT 239
QY 232 LSNNKFLYFEGNAKATT---GGAICNTKASGSPPELLISNNKTLFPASVAETSGAIAAK 288
DB 240 ITGMTSLVFSNSVYATATNGGAL-----SGDADVTISGNSVTFSGQAIVANGAIAIAK 294
QY 289 KLAASS--GGTFEFLRNVSATP-KGAISIDASGELSLSAEGNTLTFVANTLTITST 345
DB 295 KLTASAGGGGGSFSSNNVQGTAGNGAISIILAEGBLSAEGADITFNNAIYAT-TP 353
QY 346 DTPKRNAINISNGKFTFLRAKKNHTIFFYDPTISE--GTSSDVLIKINGSAGALNPYG 403
DB 354 QTKRNSIDISTAKITNLRAISGHSIFFYDPTITANTADSTDTLNLKADAGNSTDVS 413
QY 404 TILFSGETLTADDELKLVADNLKSFQPVLSGKLLQKGYLTLESTFSQEGSILGMS 463
DB 414 SIVFSGEKLSEDEAKVADNLSTLTKOPVTLTLAGNLVLRKGYTLDTKGFQTAGSSVIMDA 473
QY 464 GTTSLTSGSTITNLGIVNDSLGKOPVSLTAKGASKVIVSGKLNIDIEGNTYESHM 523
DB 474 GTTLKASTEEVTLGLSIPVDSLEGKRVIAASAKSNVALSGPILLDNOGNAYENHD 533
QY 524 FSHQFLSLKITVDADVDTNVDISSLIPVADDPENSEYFGOGGNVMTDTAT--WTX 581
DB 534 LGKTODSFVQLSA-LGTAITTTDVA---VPTVATPHYGTGTGTMGMVDDTASTPRTK 589
QY 582 EATATWTKTGVPSPERKSALVCNTLWGVFTDIRSLQOLVEIGATGMEHKOGFWWSMTN 641
DB 590 TATLAWNTGYLPNEROGPLVPNSLMGFSFDIAIQGVIERSALTLCSDRGFMAAGVAN 649
QY 642 FLHKTGDNKRKGFRTSGGYIIGSAHTPKDDLTFPACHLPARDKOCFLAHNSRTYGG 701
DB 650 FLDDKKGKKRRYRHSKSGYAIGGAQCSNLSIFACQLFSGSKDPLVAKNHTDYAG 709
QY 702 TLFFKHSHTLQPVNLRGRAKFSESAIEKPP---REIPLLDVQVSSHNDNREHYT 758
DB 710 AFYIQHITTEC-----SGFTICLLDKLPKGSMSHKPLVLEGLAASHVSNDLKTYT 759
QY 759 SLPSEGSMSNECIAGIGLDLPVLSNPH---LFEKFTIPQMKVENVYVSQNSFEFS 814
DB 760 AYPEVKSGWGNNAFMNMAGS-----SHSYPEYLHCFTYAPYIKLNTLYIROFSFSKG 814
QY 815 SDGFGFSTIGRLNLSIPYGAFFVGDIDSTYTLDSGFFVSDVYRNNQSTATLYMSDS 874
DB 815 TEGRFSDSNLFNLSLPIGVKFEKFSDCNDSYDLTLLSYVDLIRNDPKCTTALVISAS 874
QY 875 WKIRGNLSRQAFLLRGSSNNVYNSNCELFGHYAMELGSRRNVVDGTLRF 928
DB 875 WEYANNLARQALQVRAGSHAFSPMEVYLQPFVYVAGSSRIYNVDLGKRFQF 928

```

```

RESULT 7
G81591
polymorphic membrane protein G family CP0303 [imported] - Chlamydomonas pneumoniae (streptococcus pneumoniae pneumoniae)
C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence: revision 31-Mar-2000 #text: change 11-May-2000
C:Accession: G81591
R:Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: G81591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-928 <REA>
A:Cross-references: GB:AE002192; GB:AE002161; MID:g7189226; PIDN:AAF38160.1; PID:g718923
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0303

Query Match 37.3%; Score 1785; DB 2; Length 928;
Best Local Similarity 40.4%; Pred. No. 2.9e-93;
Matches 385; Conservative 173; Mismatches 344; Indels 52; Gaps 18;

QY 1 KTSIPVILVSSVLA--FSCHLQSLANEELSPDPSFNGNIDSGTFPPKTSAT--TYSILT 56
DB 1 MKSQSWILVLSSTLACFCSCSVFAATAMENIGSPDSFSGSINTGYTFKNTTGTIDYLT 60
QY 57 GDVFEPEGKGTPLSDSCFKQTTDNLTFLGCHSLTFGFDAGTHAGAASTYANKLTF 116
DB 61 GDTLQNLGDSALATKCGFSDTELSFAGKYSLSFNNIKSSAE-GALSYTDDKNLSL 119
QY 117 SGFSLSPDSSPTVT--GOGTSSAGVNLERILVYAGNFTSDGAIKASFL 174
DB 120 TGFSLITFLAAPSYYITTPSGKAVKCGDGLTFDNGITLTKQDYCEENGAISTRKNSL 179
QY 175 TGTSGDALFSSNNSSST---KGAIAATAGARIANTNTGYVRFSLNIASTSGAIDDEGTSI 231
DB 180 KNSGTSISFEGKSSAATCKKGALCATGTVDITNTTAPTLFSNNIAEAGAAGINSTGCT 239
QY 232 LSNNKFLYFEGNAKATT---GGAICNTKASGSPPELLISNNKTLFPASVAETSGAIAAK 288
DB 240 ITGMTSLVFSNSVYATATNGGAL-----SGDADVTISGNSVTFSGQAIVANGAIAIAK 294
QY 289 KLAASS--GGTFEFLRNVSATP-KGAISIDASGELSLSAEGNTLTFVANTLTITST 345
DB 295 KLTASAGGGGGSFSSNNVQGTAGNGAISIILAEGBLSAEGADITFNNAIYAT-TP 353
QY 346 DTPKRNAINISNGKFTFLRAKKNHTIFFYDPTISE--GTSSDVLIKINGSAGALNPYG 403
DB 354 QTKRNSIDISTAKITNLRAISGHSIFFYDPTITANTADSTDTLNLKADAGNSTDVS 413
QY 404 TILFSGETLTADDELKLVADNLKSFQPVLSGKLLQKGYLTLESTFSQEGSILGMS 463
DB 474 GTTLKASTEEVTLGLSIPVDSLEGKRVIAASAKSNVALSGPILLDNOGNAYENHD 533
QY 524 FSHQFLSLKITVDADVDTNVDISSLIPVADDPENSEYFGOGGNVMTDTAT--WTX 581
DB 534 LGKTODSFVQLSA-LGTAITTTDVA---VPTVATPHYGTGTGTMGMVDDTASTPRTK 589
QY 582 EATATWTKTGVPSPERKSALVCNTLWGVFTDIRSLQOLVEIGATGMEHKOGFWWSMTN 641
DB 590 TATLAWNTGYLPNEROGPLVPNSLMGFSFDIAIQGVIERSALTLCSDRGFMAAGVAN 649
QY 642 FLHKTGDNKRKGFRTSGGYIIGSAHTPKDDLTFPACHLPARDKOCFLAHNSRTYGG 701
DB 650 FLDDKKGKKRRYRHSKSGYAIGGAQCSNLSIFACQLFSGSKDPLVAKNHTDYAG 709
QY 702 TLFFKHSHTLQPVNLRGRAKFSESAIEKPP---REIPLLDVQVSSHNDNREHYT 758
DB 710 AFYIQHITTEC-----SGFTICLLDKLPKGSMSHKPLVLEGLAASHVSNDLKTYT 759
QY 759 SLPSEGSMSNECIAGIGLDLPVLSNPH---LFEKFTIPQMKVENVYVSQNSFEFS 814
DB 760 AYPEVKSGWGNNAFMNMAGS-----SHSYPEYLHCFTYAPYIKLNTLYIROFSFSKG 814
QY 815 SDGFGFSTIGRLNLSIPYGAFFVGDIDSTYTLDSGFFVSDVYRNNQSTATLYMSDS 874
DB 815 TEGRFSDSNLFNLSLPIGVKFEKFSDCNDSYDLTLLSYVDLIRNDPKCTTALVISAS 874
QY 875 WKIRGNLSRQAFLLRGSSNNVYNSNCELFGHYAMELGSRRNVVDGTLRF 928
DB 875 WEYANNLARQALQVRAGSHAFSPMEVYLQPFVYVAGSSRIYNVDLGKRFQF 928

```

[illegible][illegible]

Query Match 100.0%; Score 4782; DB 2; Length 949;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-263;  
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKTSPWLVSVSLAFSCHLQSLANEELLSPDDSFNGNIDSGTPTPKTSATYSLTGDFV 60  
 22 MKTSPWLVSVSLAFSCHLQSLANEELLSPDDSFNGNIDSGTPTPKTSATYSLTGDFV 81  
 61 FYEPEGKPLSDSCFKQTTDNLTFLGNCHSLTFPGTIDAGTHAGAASTTANKNLTFPGS 120  
 82 FYEPEGKPLSDSCFKQTTDNLTFLGNCHSLTFPGTIDAGTHAGAASTTANKNLTFPGS 141  
 121 LLSPPSSPTVTGTGGGTLSSAGVNLNIRKLVAGNPTSDGATIGASFLTLGTSGD 180  
 142 LLSPPSSPTVTGTGGGTLSSAGVNLNIRKLVAGNPTSDGATIGASFLTLGTSGD 201  
 181 ALFNSNSSSTKGAIATAGARIANNNTGYVRLSNIASTSGAIDEGTSLSNKKFLYF 240  
 202 ALFNSNSSSTKGAIATAGARIANNNTGYVRLSNIASTSGAIDEGTSLSNKKFLYF 261  
 241 EGNAAKTGGAICNTKASGSPELLISNNKTLIFASNAVETSOGAIIHAKKLAISGCTFE 300  
 262 EGNAAKTGGAICNTKASGSPELLISNNKTLIFASNAVETSOGAIIHAKKLAISGCTFE 321  
 301 LRNNVSSATPKGAIISIDASGELSLAETGNTTFVRNTLTGSGTDPKRNAINIGSNGK 360  
 322 LRNNVSSATPKGAIISIDASGELSLAETGNTTFVRNTLTGSGTDPKRNAINIGSNGK 381  
 361 FTFLRAAKNHTIFFYDPTLSEGTSSDVLIKINNGSAGALNPYOGTLLFSGETTLADELVA 420  
 382 FTFLRAAKNHTIFFYDPTLSEGTSSDVLIKINNGSAGALNPYOGTLLFSGETTLADELVA 441  
 421 DNKSSFPQPVSLSGKLLQKGVLTLESTSPSQEAGSLGMDSGTTLSTAGSTITNIG 480  
 442 DNKSSFPQPVSLSGKLLQKGVLTLESTSPSQEAGSLGMDSGTTLSTAGSTITNIG 501  
 481 INVDSLGKOPVSLTAKASNKVYISGKLNLDIEGNTYESHMFHDLFSLLKTLTVAD 540  
 502 INVDSLGKOPVSLTAKASNKVYISGKLNLDIEGNTYESHMFHDLFSLLKTLTVAD 561  
 541 VDTNVDISSLIPVAEDPENSEYGFQGMNVMWTTDTATNTKEATATWTKTGFVSPERKS 600  
 562 VDTNVDISSLIPVAEDPENSEYGFQGMNVMWTTDTATNTKEATATWTKTGFVSPERKS 621  
 601 ALVONTLMGVTTDIRSLQQLVEIGATGMEHQGFVSSMTNFLHKTGDBNKGFFHTSG 660  
 622 ALVONTLMGVTTDIRSLQQLVEIGATGMEHQGFVSSMTNFLHKTGDBNKGFFHTSG 681  
 661 YVIGSATHPRDGLFTFAFCHLFAADKCFIAHNNSTRYGGTLFFKHSHTLQPOVYLRG 720  
 682 YVIGSATHPRDGLFTFAFCHLFAADKCFIAHNNSTRYGGTLFFKHSHTLQPOVYLRG 741  
 721 RAKFSESAIEKFPREIPLADVOVSFHSNDNMETHYTSLEPSEGSWSNECTIAGIGDL 780  
 742 RAKFSESAIEKFPREIPLADVOVSFHSNDNMETHYTSLEPSEGSWSNECTIAGIGDL 801  
 781 PFVLSNPRLKFTFIPQKVMVYVYSSONSFESSSDGRGFSIGRLNLSIPVGAFFVGD 840  
 802 PFVLSNPRLKFTFIPQKVMVYVYSSONSFESSSDGRGFSIGRLNLSIPVGAFFVGD 861  
 841 IGDVYTDLSCGFVSDVYRNNPOSTATLWSPDSMKIRGNSLSROAFLIRGSNNVYSN 900  
 862 IGDVYTDLSCGFVSDVYRNNPOSTATLWSPDSMKIRGNSLSROAFLIRGSNNVYSN 921  
 901 CELFGHAMELRGSSRNYNVDVGTKLRF 928  
 922 CELFGHAMELRGSSRNYNVDVGTKLRF 949

RESULT 4  
 E86546  
 polymorphic outer membrane protein G/I family (imported) - Chlamydia pneumoniae (str

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
 C:Accession: E86546  
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A:Reference number: A86491; MUID:20330349  
 A:Accession: E86546  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-928 <STO>  
 A:Cross-references: GB:BA000008; NID:98978819; PID:BA098655.1; GSPDB:GN00142  
 A:Experimental source: strain J138  
 C:Genetics:  
 A:Gene: pmp\_9

Query Match 37.5%; Score 1794; DB 2; Length 928;  
 Best Local Similarity 42.7%; Pred. No. 9.1e-94;  
 Matches 404; Conservative 152; Mismatches 353; Indels 38; Gaps 17;

1 MKTSPWLVSVSLAFSCHLQ-----SLANEELLSPDDSFNGNIDSGTPTP-----KTSAT 51  
 1 MKTSPWLVSVSLAFSCHLQ-----SLANEELLSPDDSFNGNIDSGTPTP-----KTSAT 57  
 52 TYSLFGDVEFYEPKGTPLSDSCFKQTTDNLTFLGNCHSLTFPGTIDAGTHAGAASTTAN 111  
 58 IYNLGDVSTNAGSPALITASCFTETGNLSFGCHGFOFLQNDADAGANC-TFTNTAAN 116  
 112 KNLTFSGSLSLSDSPSTVTYTGGLTSSAGVNLNIRKLVAGNPTSDGAIKIGAS 171  
 117 KLSFGSGLYSL--IQTNATGTGAIKSTGACISQVSCYFGQNFNSNDGALQSS 174  
 172 FLTLTSGDALFNSNSSSTKGAIATAGARIANNNTGYVRLSNIASTSGAIDDEGSI 231  
 175 ISLS-LNPNLTFKAKKATOKGALYSTGGITINNTLNASFSEMTAANNNGAIYTEASSF 233  
 232 LSNNFLYFEGN--AAATGGAI-CNTKASGPELLISNNKTLIFASNAVETSOGAIIHA 287  
 234 ISSNAIFINSVYATATGATGALYCSTSAKPLVLTSDNGELNFTGNTATTSOGALYT 293  
 288 KKLASSGCTFEFLNN--VSSATPKGAIISIDASGELSLAETGNTTFVRNTLT--TGST 345  
 294 DNVLVSGGPTLFKNNSAIDTAAPLGAIALADSGSLSLAIGDITPEGNTVYVAGASS 353  
 346 DTPKRNAINIG-SNKGTEFLRAKKNHTFFYDPTLSEGTSSDVLIKINNGSAGALNPY 402  
 354 QTTNRSINIGTNAKIVOLRASQGTTFYDPTLSTTALSDALNLNGPDLAGNPAYO 413  
 403 GTLFSGFTLFADELKADNLSKSTQVPSLGGKLLQKGVLTLESTSPSQEAGSLGMD 462  
 414 GTIVSGERLSAEAEADNLSKSTIQOPLTLAAGLSLKSQVTLVAKFSQSPGSTLMD 473  
 463 SGTTLSTAGSTITNIGLINVDSLGKOPVSLTAKASNKVYISGKLNLDIEGNTYESH 522  
 474 AGTLETETADG--ITINNVLNVDLSKETKATLAKAQASQVTLSSLSLVPSGNNVEDV 532  
 523 MFSHQLSLKLTIDAVDNRVNDISLIPVAEDPENSEYGFQGMNVMWTTDTATNTKE 582  
 533 SWNNQVPSCLTLT--ADDPANIHITDLAADLENNPLHWGQGMWALSMOEDATATKKA 590  
 583 ATATWTKTGFVSPERKALVNTLMGVTTDIRSLQQLVEIGATGMEHQGFVSSMTNF 642  
 591 ATLTTWTKTGFVSPERKALVNTLMGVTTDIRSLQQLVEIGATGMEHQGFVSSMTNF 650  
 643 LKHTGDBNKGFFHTSGVYVIGSATHPRDGLFTFAFCHLFAADKCFIAHNNSTRYGGT 702  
 651 FHKDSTKLNKGRHISAGYVYGATTTLASDNLITAFQOLFGKDRHFINNKRAVAAAS 710  
 703 LFFKHSHTLQPOVYLRKRAKFSESAIEKFPREIPLADVOVSFHSNDNMETHYTSLE 762  
 711 LHLQHLATLSSPSLRY--LPGSES-----EQVLPDAQISYISKNTKTYTTOAPK 761

**RESULT**

**2**

D7077  
polymorphic outer membrane protein g family - Chlamydomophila pneumoniae (strain CWL029)  
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: D72077  
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.  
Nature Genet. 21, 385-389, 1999  
A:title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: D72077  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-928 <ARN>  
A:Cross-references: GB:AEO01628; GB:AEO01363; NID:g4376730; PIDN:AND18593.1; PID:g437673  
C:Genetics:  
.:gene: pmp\_11

Query Match	100.0%	Score 4782	DB 2	Length 928
Best Local Similarity	100.0%	Pred. No. 9,4e+263		
Matches	928	Conservative	0	Mismatches 0; Indels 0; Gaps 0;
QY	1	MKTSIPWLVSSVLAFCSHLOSLANEELLSPDPSFNGNIDSGTFEPTPKTSATYSITGDFV	60	
Db	1	MKTSIPWLVSSVLAFCSHLOSLANEELLSPDPSFNGNIDSGTFEPTPKTSATYSITGDFV	60	
QY	61	FYEFGKGPILSDSCFKQTTDMLTFLGNGHSLTFEFDIDAGTHAGAAASTTANKNTLTFSGFS	120	
Db	61	FYEFGKGPILSDSCFKQTTDMLTFLGNGHSLTFEFDIDAGTHAGAAASTTANKNTLTFSGFS	120	
QY	121	LISPDSSSTVTYTGOGGTLSSAGAGVNLINIRKLVAAGNFSTADGATIKGASFTLLTGTSGD	180	
Db	121	LISPDSSSTVTYTGOGGTLSSAGAGVNLINIRKLVAAGNFSTADGATIKGASFTLLTGTSGD	180	

QY	181	ALFNNSSSTGGAATATAGARIANNCTYVPLSIASSTSGAIDDEBCTSLSNKKPLXF	240
Db	181	ALFNNSSSTGGAATATAGARIANNCTYVPLSIASSTSGAIDDEBCTSLSNKKPLXF	240
QY	241	EGNAKTTGGAICNTKASGPELIIISNNKTLIFASNVETSGAIIHAKKLLSSGCFEEF	300
Db	241	EGNAKTTGGAICNTKASGPELIIISNNKTLIFASNVETSGAIIHAKKLLSSGCFEEF	300
QY	301	LRNNVSATPKGAGISIDASELSIAETGNTFVRNTLTTTGTSTDPKRNAINIGSNGK	360
Db	301	LRNNVSATPKGAGISIDASELSIAETGNTFVRNTLTTTGTSTDPKRNAINIGSNGK	360
QY	361	FTELRAAANHITFFPDPTTSSGTSDDVAKINNGSAGALNPQGTILPSGELTFADELKVA	420
Db	361	FTELRAAANHITFFPDPTTSSGTSDDVAKINNGSAGALNPQGTILPSGELTFADELKVA	420
QY	421	DNLKSSFPQPSLSGGKLLQKGYLTESTSFQSEAGSLGMDSGTTLSTTGSTITYNLG	480
Db	421	DNLKSSFPQPSLSGGKLLQKGYLTESTSFQSEAGSLGMDSGTTLSTTGSTITYNLG	480
QY	481	INVDSLGKQPVSLTAKGASNKYIVSGKLLIDIEGNIYESHMFSDQLFSLKITVDAD	540
Db	481	INVDSLGKQPVSLTAKGASNKYIVSGKLLIDIEGNIYESHMFSDQLFSLKITVDAD	540
QY	541	VDITNDISSLPVAEDPNSEYGFQGGOMNVMITDITANTKEAATATKTKTFVSPERKS	600
Db	541	VDITNDISSLPVAEDPNSEYGFQGGOMNVMITDITANTKEAATATKTKTFVSPERKS	600
QY	601	ALVCTTLGVTDTIRSLQOLVEIGATGMEHKQGFVWSSMTNFKLTGDENRKGRHISGG	660
Db	601	ALVCTTLGVTDTIRSLQOLVEIGATGMEHKQGFVWSSMTNFKLTGDENRKGRHISGG	660
QY	661	YVIGGSAHTPKDDLFTFAFCHLFARDKDCFIAHNSRTYGGTLEFKHSHTLOPONYLRIG	720
Db	661	YVIGGSAHTPKDDLFTFAFCHLFARDKDCFIAHNSRTYGGTLEFKHSHTLOPONYLRIG	720
QY	721	RAKSESIAEKFPEEIPALIDVOYSESHSDRMETHYSLPESEGSWNBEICIAIGIGIDL	780
Db	721	RAKSESIAEKFPEEIPALIDVOYSESHSDRMETHYSLPESEGSWNBEICIAIGIGIDL	780
QY	781	PFLVLSNPAPLPEKTFIPQKVMEMVYVQNSFFESSSDGKGFSGIRLLNLISIPVGAKFYOGD	840
Db	781	PFLVLSNPAPLPEKTFIPQKVMEMVYVQNSFFESSSDGKGFSGIRLLNLISIPVGAKFYOGD	840
QY	841	IGDSYTYDLSGFVYSDVYRNNPOSTATILVMSPDMSKIRGNLSQOAFLLRGSNNYVNSN	900
Db	841	IGDSYTYDLSGFVYSDVYRNNPOSTATILVMSPDMSKIRGNLSQOAFLLRGSNNYVNSN	900
QY	901	CELFGHAMELRGSSRNYNVDVGTKLRF 928	
Db	901	CELFGHAMELRGSSRNYNVDVGTKLRF 928	

RESULT 3  
F81591  
polymorphic membrane protein G family CP0302 [imported] - Chlamydomonas pneumoniae (s  
C:Species: Chlamydomonas pneumoniae, Chlamydomonas pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: F81591  
C:Read, T.D.; Brnham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis Morn and Chlamydia pneumoniae AR39  
A:Reference number: AB1500; MUID:20150255  
A:Accession: F81591  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-949 <REA>  
A:Cross-references: GB:AE002192; GB:AE002161; NID:g7189226; PIDN:AAF38159.1; PID:g718  
A:Experimental source: strain AR39, HL cells  
C:Genetics;  
A:Gene: CP0302

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: October 2, 2001, 03:07:03 ; Search time 76.79 Seconds

(without alignments)  
920.362 Million cell updates/sec

Title: US-09-446-677b-2

Perfect score: 4782

Sequence: 1 MKTSIPWLVSVLAFSCHL.....MELRGSSRNRYVDVGTFLRF 928

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Database : PIR\_68:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4782	100.0	928	2	H86546 polymorphic outer
2	4782	100.0	928	2	H86546 polymorphic outer
3	4782	100.0	949	2	F81591 polymorphic membra
4	1794	37.5	928	2	E86546 polymorphic outer
5	1794	37.5	928	2	B72077 polymorphic membra
6	1785	37.3	928	2	G86546 polymorphic outer
7	1785	37.3	928	2	G81591 polymorphic membra
8	1763	36.9	930	2	D86546 polymorphic outer
9	1763	36.9	930	2	A81591 polymorphic membra
10	1761	36.8	930	2	D72078 polymorphic outer
11	1734	36.3	936	2	C72077 polymorphic outer
12	1733	36.2	936	2	C86546 polymorphic outer
13	1733	36.2	936	2	B81591 polymorphic membra
14	1591	33.3	772	2	H86492 Pmp_3 [imported] -
15	1435.5	30.0	841	2	E72130 polymorphic membra
16	1353	28.3	922	2	B72131 polymorphic outer
17	1353	28.3	922	2	E86491 polymorphic outer
18	1352	28.3	922	2	F81539 polymorphic membra
19	1284.5	26.9	973	2	B86547 polymorphic outer
20	1284.5	26.9	973	2	F72076 polymorphic membra
21	1284.5	26.9	995	2	C81593 polymorphic outer
22	1203	25.2	1276	2	B86546 polymorphic membra
23	1203	25.2	1276	2	C81591 polymorphic outer
24	1164	24.3	712	2	E86492 polymorphic membra
25	1144.5	23.9	1407	2	G71460 polymorphic outer
26	1102.5	23.1	1013	2	G71460 probable outer mem
27	1040	21.7	987	2	H81722 polymorphic membra
28	843.5	17.6	494	2	D86493 polymorphic outer
29	837	17.5	878	2	B71460 probable outer mem

30	831.5	17.4	427	2	A86493 polymorphic outer
31	829.5	17.3	867	2	F81721 polymorphic membra
32	790	16.5	445	2	E86493 Pmp_5 [imported] -
33	736.5	15.4	359	2	C86493 Pmp_4 [imported] -
34	679.5	14.2	947	2	D72067 polymorphic membra
35	678.5	14.2	947	2	G86557 polymorphic membra
36	610.5	12.8	964	2	E71460 polymorphic outer mem
37	610.5	12.8	1723	2	H86557 polymorphic membra
38	610.5	12.8	1723	2	E72067 polymorphic membra
39	610.5	12.8	1732	2	C81601 polymorphic membra
40	608.5	12.7	978	2	B81593 polymorphic membra
41	608.5	12.7	978	2	C86547 polymorphic outer
42	604.5	12.6	986	2	B81675 polymorphic membra
43	604	12.6	1016	2	H71460 probable outer mem
44	603.5	12.6	978	2	G72076 polymorphic outer
45	600.5	12.6	934	2	G86548 polymorphic outer

## ALIGNMENTS

RESULT 1

H86546

polymorphic outer membrane protein G family [imported] - Chlamydomonas pneumoniae (st

C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001

C:Accession: H86546

R:Shitai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;

Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349

A:Accession: H86546

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-928 &lt;STO&gt;

A:Cross-references: GB:BA000008; NID:98978822; PIDD:BA98658.1; GSPDB:GND00142

A:Experimental source: strain J138

A:Genetics:

A:Gene: pmp\_11

Query Match	100.0%	Score 4782	DB 2	Length 928
Best Local Similarity	100.0%	Pred. No. 9.4e-263		
Matches 928	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MKTSIPWLVSVLAFSCHLQSLANEELSPDSEFGNIDSGFTPKSATYSLGDF	60	
DB	1	MKTSIPWLVSVLAFSCHLQSLANEELSPDSEFGNIDSGFTPKSATYSLGDF	60	
QY	61	FYEPGKGTPLSDSCFEKQTDNLTFLGNGHSLTFGFIAGTHAGAAATYNNLTFSGFS	120	
DB	61	FYEPGKGTPLSDSCFEKQTDNLTFLGNGHSLTFGFIAGTHAGAAATYNNLTFSGFS	120	
QY	121	LTSFDSPTTGTGGTLLSAGVNLNIRKLVVAGNFSTADGCAIKASFLTGTSDD	180	
DB	121	LTSFDSPTTGTGGTLLSAGVNLNIRKLVVAGNFSTADGCAIKASFLTGTSDD	180	
QY	181	ALFSNNSSSTKGATATTGARIANTGVVRLSNASTSGAIDDEGSIISNNKELF	240	
DB	181	ALFSNNSSSTKGATATTGARIANTGVVRLSNASTSGAIDDEGSIISNNKELF	240	
QY	241	EGNAKTGTGALCNFKASGSPPELLISNNKTLIFASVNAETSGAIIHAKKIALSSGFTF	300	
DB	241	EGNAKTGTGALCNFKASGSPPELLISNNKTLIFASVNAETSGAIIHAKKIALSSGFTF	300	
QY	301	LNNVSSATPKGALSIDASGELSIAETGNTFVARNLTGTSDDTPKRNAINISNGK	360	
DB	301	LNNVSSATPKGALSIDASGELSIAETGNTFVARNLTGTSDDTPKRNAINISNGK	360	
QY	361	FTEIRAKNHTTFYDPTISEGSSDYLTKINNGSAGALPYGTLIFSETTLADELKYA	420	
DB	361	FTEIRAKNHTTFYDPTISEGSSDYLTKINNGSAGALPYGTLIFSETTLADELKYA	420	

**THIS PAGE BLANK (USPTO)**



**BEST AVAILABLE COPY**

---



```

Db 121 SGFSLFSCAYCCPCT----- 136
Qy 175 TGTSDALFNNSSSTKGAATATTAGARIANTGVRLSNASTSGAIDDEGSIISN 234
Db 137 -----TGY-----GAIQTKGTTTLKD 152
Qy 235 NKFLYFEGNAKTTGCAICNTRKASGPPELLISNNKTLIFASNAVETSGAIAHAKLALS 294
Db 153 NSSLVFHNKCSAEBGALQCKSSSTALKEKNKLVFSESSKREKGAIAADLTIYS 212
Qy 295 GGFEEFLNNVS-SATPKGALIS-DASGEJLSAETGNITFVRN-TLTTGSTDTPKRN 351
Db 213 GGPFLFSNNSSVSHNSPGGALICIDSDGECSLFANLGDITFDGKIIITNGSGPTVRN 272
Qy 352 AINNGSKFTELRAKHNHTFFYDPIRSESSDYLNKINNSAALNPGCTILFSET 411
Db 273 SIDGSGKFTKLNKEGFGFEFTPIANTGGSTET-ELNKTESD--TYYGKIYFSEK 329
Qy 412 LTADLKVADNLKSSFTQPVSLSGKLLLOKGVTLTESTFSQEAAGSLGMDGTTLSTTA 471
Db 330 LSDEKATVPANLKSFKPKIKIGAGSLVLDKGVTLFAKKIKYTKSGTYVMDLGTTLQTPS 389
Qy 472 GS---TITNIGINDSI--GLKQVSLAKGASNKVYVSGKLNLDIEGNIYESHMS 525
Db 390 SSGETITLTINDINIASIGGGGTPAKLATMTASQAISIAA-VNLVNTDSNTYEDPILS 448
Qy 526 HDQFLSLKIVDADVDPNVNDISLIPVPAED-----PNSGPGOGNNVMTDTATNT 580
Db 449 ASKSSAATTAATTS-----SSTVTPETLNKNTPTTHGYQGNNTVTKGSSAOE 499
Qy 581 KEATATWKTGFVSPERKSALVNTLNGVFTDIRSLQOLVEIGATGMEHKOGFVSSMT 640
Db 500 KTAITLTWQOTGSPRPERVGSIVPTLNGAFSDPTAIONLMDISVNGADYSKGFVSSLA 559
Qy 641 NPLHKTGDENNKGERHRTSGGYVIGSAHTPKDILTFAPCHLPARDKCFIAHNSRTYG 700
Db 560 NPLNKSAGDTRKFRKHHAGYALGYAQTPSDYCASACOLFGKDKDYFVSKNSTIYA 619
Qy 701 GTLPFKHSHTLQOPNYL--RLGRAKFSESALIEKPRRIPLADVOVFSHSDNMEHY 757
Db 620 GSIIYQHSHTYNTWTTLQNTLG-----AEPALVLNQLVYCHASNNMKNTM 666
Qy 758 TSL-----PESGWSNECIAAGIGLDPFVLNPHLPFKPTIPOMKEMVYVSONS 809
Db 667 TINTTPKRVNTSEIKGMDGNDGFCGVEBKAAP--IETASLLEFDMSPVKQLVIAHDD 724
Qy 810 FFESSD--GRGFSIGRLNLSIPVGAFTVQDIDGSDTYTDLSGFFVSDVYRNNPOSTATL 868
Db 725 FKENNSDQGRFESNNLNLNLSMPLEKFSHKDTASYNLTLAVAPDIVRSNPCTASL 784
Qy 869 VMSDPS--WKIRGNLSQAFLLRGSNNYVNSNCELFCHAMELRGSSRYNNVDTGKL 926
Db 785 LVSPSAVWTKANLHAHAFILQAGNYLALTRNTEFSQFELRGSCRYNIDLGSKI 844
Qy 927 RF 928
Db 845 QF 846

```

```

Rx MEDLINE-96406378; PubMed-8810511;
Ra Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
Rt Identification of a multigene family coding for the 90 kda proteins
Rt of the ovine abortion subtype of Chlamydia psittaci.
Rn FEMS Microbiol. Lett. 142:277-281(1996).
Rn [2]
Rp SEQUENCE FROM N.A.
Rc STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
Rx MEDLINE-98187897; PubMed-9529048;
Ra Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
Rt Molecular cloning and characterization of the genes coding for the
Rt highly immunogenic cluster of 90-kilodalton envelope proteins from the
Rt Chlamydia psittaci subtype that causes abortion in sheep.
Rl Infect. Immun. 66:1317-1324(1998).
Dr EMBL: U65942; AAC15921.1; -.
Dr InterPro: IPR003357; -.
Dr InterPro: IPR003368; -.
Dr Pfam: PF02385; OMP; 1.
Dr Pfam: PF02415; DUF145; 1.
Dr SEQUENCE 847 AA; 90694 MW; 754C958E7F1179E CRC64;

Query Match 32.0%; Score 1532.5; DB 2; Length 847;
Best Local Similarity 37.0%; Pred. No. 1.9e-82;
Matches 357; Conservative 150; Mismatches 302; Indels 157; Gaps 23;

Qy 1 MKTSIPWLVSSVLAFSCHLQSIANEE---LSPDSEFNKINISGTFP---KTSATF 52
Db 1 MKNPVMFLISSLSIVS--NSLYSEPDQKTLVSAHSYNNNTNNEPNNPSTSNISNCT 57
Qy 53 YSLNGDVFVEYEPG-KGTPLSDSCEKOTYDNLFLNGHSLTFGFIDA-GTHAGAASTYA 110
Db 58 YTCGNICIAAAGLDGSSLSGSCFTDAGNLSFLGNGYTLCPDNTTQSHPGAISVSGT 117
Qy 111 NKNLTSGFSLISFSDSPSTVYTTGQTLSSAGVNLNIRKLVYAGNFSTADGAIKGA 170
Db 118 NKTLDISGFSLSFCAYCP----- 135
Qy 171 SPLLTGTGDALFNNSSSTKGAATATTAGARIANTGVRLSNASTSGAIDDEGTS 230
Db 136 -----PGA---TGY-----GAIKAVGNT 150
Qy 231 ILSNKKFLYFEGNAKTTGCAICNTRKASGPPELLISNNKTLIFASNAVETSGAIAHAKL 290
Db 151 TIKDNSSLVFHKNCSTGCGAIOCKASSAELEKKNENNLVFAENSSSSSGAIAADKL 210
Qy 291 ALSSGFTPELRNNVVSATPKGALIS-DASGEJLSAETGNITFVRN-TLTTGSTDTP 348
Db 211 TIVSGGPFLFSNNSSVSHNSPGGALICIDSDGECSLFADLGDITFDGKIIITNGSGPTV 270
Qy 349 KRNAINIGSKFTELRAKHNHTFFYDPIRSESSDYLNKINNSAALNPGCTILFSES 408
Db 271 TRNSIDLGSSGKFTKLNKEGFGFEFTPI--GGGSELNINKODT---VDYTKIYFES 325
Qy 409 GETLTADLKVADNLKSSFTQPVSLSGKLLLOKGVTLTESTFSQEAAGSLGMDGTTLST 468
Db 326 GERLSDEKKAANLKSDFKQPKIKIGAGSLVLDKGVTLFAKKIKYTKSGTYVMDLGTTLQ 385
Qy 469 TTAG---TITNIGINDSI--LKQPVSLAKGASNKVYVSGKLNLDIEGNIYES 521
Db 386 TPSSGGETITLTINDINVASLGGGVADPAKVEATTESKTVTINA-VNLVDDGNVAYE 444
Qy 522 HMFSDQFLSLKITYDADVDPNVNDISLIPVPAED-----PNSGPGOGNNVMTDT 576
Db 445 PTLASQPFYTAIEVRSGS-----SGSITKPTTNLENYTPPHYTGQGMWYTWKQGS 496
Qy 577 AATNKEATATWKTGFVSPERKSALVNTLNGVFTDIRSLQOLVEIGATGMEHKOGFV 636
Db 497 SAOKEATALTWQOTGSPRPERVGSIVPTLNGAFSDPTAIONLMDISVNGADYHRGFW 556
Qy 637 SSMNTFLHKTGDENNKGERHRTSGGYVIGSAHTPKDILTFAPCHLPARDKCFIAHNS 696
Db 557 SGLGNFLHKSQDTRKFRHNSAGYALGYAQTPESEDFSAACOLFGKDKDYLVSKNS 616

```

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. [1]  
 OX NCBI\_taxid=83554;  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;  
 RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;  
 RL Submitted (Oct-1996) to the EMBL/Genbank/DBD databases.  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;  
 RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;  
 RL Submitted (Oct-1996) to the EMBL/Genbank/DBD databases.  
 DR EMBL; U05943; AAC15924.1; -  
 DR EMBL; U05942; AAC15922.1; -  
 DR InterPro; IPR003357; -  
 DR InterPro; IPR003368; -  
 DR Pfam; PF02385; OMP; 1.  
 DR Pfam; PF02415; DUF145; 1.  
 DR Signal.  
 KW Signal.  
 FT CHAIN 17 839 POMP90B.  
 SO SQUENCE 839 AA; 89824 MW; 4581C7CBA7EFAC4 CRC64;

Query Match 32.7%; Score 1562.5; DB 2; Length 839;  
 Best Local Similarity 37.7%; Pred. No. 3.1e-84;  
 Matches 362; Conservative 156; Mismatches 290; Indels 151; Gaps 26;

OY 1 MKTSIPWVSSVLAFSCHQSLAN--ELLSPDDSFNGNIDSGTFPRK--TSATYSLT 56  
 DB 1 MKHPVWFLLISSSL-FASNSLSFANAOATALTSPDSNGVNTSEERQVKTSSGTYTCE 59  
 OY 57 GDFEYEPGKGTPLSDSCFKOTDNLTLFNGHSLTFEGFI-DAGTHAGAASTYANKLTF 115  
 DB 60 GNVCFISAGAGDSGLKSCF-SATDNLTLFNGHSLTFEGFI-DAGTHAGAASTYANKLTF 118  
 OY 116 FSGPSLSDPSSTVYTTGGTSLSSAGVNLNIRKLVVAGNFSTADGAIKASFLLT 175  
 DB 119 ISGFSLSFCAYCPGPT----- 134  
 OY 176 GTSGDALFSNNSSTKGATATTAGARIANNVYRFLNSIASTSGAIDDEGTSLISNN 235  
 DB 135 -----TGY-----GAIOTKKNNTLKD 151  
 OY 236 KFLYFEGNAKTTGATACNTKAGSPPELLISNNKTLIFASNAVETSGAIIHAKKIALSSG 295  
 DB 152 SSVLFHKNCTAGAGAT-OCKGSSDAELKLENNONLVFSENSSTSGGAIYADKLIVSG 210  
 OY 296 GTFEFLANNVS-ATPKGAISI-DASGELSLAETGNITFVRN-FLTWTGSTDTPKRNA 352  
 DB 211 GPFLFSNNSVSGSPKGAISIKDSSGECSLADGDTTFDGNKIIKTSGGSGTYTRNS 270  
 OY 353 INGSNGKTEFLAAKNHTFEFYDPTTSEGTSSDYKINNKGAGALNPGDTLFGSEGL 412  
 DB 271 IDLGT-GKFTKFLAKAGGFGFFYDPTT--GGGSDDELINKKET--VDYGTKIVFSGEKL 324  
 OY 413 TADELVADNLKSFQPSYLSGKLLQKGYLLESTSFSGQSLGMLSGTLLSTAG 472  
 DB 325 SDEKARAEMLASTFNQPTLSAGSLVKDGYVTAQYQEGSYVMGLGTLTLPSS 384  
 OY 473 ---SITTNGLINVDL--GLKQPVSLAKGASNKVYSGKLNLDIEGNIYSHMFH 536  
 DB 385 GGEITLTINDIMIASLGGGGTSPAKLATNTASQAITTA--VNLVADGNAVEDPIIAT 443  
 OY 527 DQFSLIKIVDADVDPNDISLIPVPAED-----PNSRYGQGMVWMTTDTATNKR 581  
 DB 444 SKFTLVAVATTA-----STVQPTDNLNTPVPHYHGQGMVYTWDETAKT- 493  
 OY 582 EATATYTKTGTVSPERKALVNTLWGVETDIRSLQQLVEIGATGMEHKGFWVSSMTN 641  
 DB 494 -ATLTWQGTGSPNPERQGLVYNTLWGAFLSDLRATQNLMDISVNGADYHARGWVGSLAN 552  
 OY 642 FLAKTGDENRKGRHTSGVYVIGSAHTPRDLFTFAFCHLFAKDKCFIAHNSRTYGG 701

DB 553 FLHKSQSDTKRKRHHNSAGALGYAKTPSDIDFSAFCOLFGKDKRXYLVSKNNANIYAG 612  
 OY 702 TLEFKH-SHTLQONTLRLGRARFSESAIEKPREIPLADVQVSPSHNRHETH--- 757  
 DB 613 SLTYQHTSYSAMONLLQ-----NTIGAEALVNAQULTYCHASNDKNTNTTTT 661  
 OY 758 ----TSLPSEGSWSNECIAGIGLDLPFLVSNPHLPFTFIPOMKEVAVYVSQNSFFE 812  
 DB 662 YARKTLYATIKDMDGDCGVELGATVP-IQTESLLFPMYSPLKFLQVLHHDQDFKE 720  
 OY 813 SSSD-GRGFSIGRLNLSTIVGAKFYVQGDIGSYTYDLSCFVSVSVYRNNPOSTATLVMS 871  
 DB 721 NNSDQGRYFESSNLTNLTSLPDIKIFERFANNDPASYHVTAASPDIVRSNPDCTTSLVS 780  
 OY 872 PDS--WKIRGNLSRQAFLLRGSNNVYNSNCELFHYANMELGSSNNYVVDGTRKRF 928  
 DB 781 PDSAVVTKANNLARSFMLQAGNYLSLSHNIEIFSOFGELRGSSRTYVNDLGSKIQF 839

RESULT 13  
 ID P71133 PRELIMINARY; PRT; 846 AA.

AC P71133;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE POMP91B PRECURSOR  
 OS Chlamydia psittaci (Chlamydia psittaci).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_taxid=83554;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;  
 RX MEDLINE=96406378; PubMed=8810511;  
 RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;  
 RT Identification of a multigene family coding for the 90 kDa proteins  
 RL of the ovine abortion subtype of Chlamydia psittaci.";  
 RL FEMS Microbiol. Lett. 142:277-281(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;  
 RX MEDLINE=98187897; PubMed=9529048;  
 RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;  
 RT Molecular cloning and characterization of the genes coding for the  
 RT highly immunogenic cluster of 90-kilodalton envelope proteins from the  
 RT Chlamydia psittaci subtype that causes abortion in sheep.";  
 RL Infect. Immun. 66:1317-1324(1998).  
 DR EMBL; U05943; AAC15923.1; -  
 DR InterPro; IPR003357; -  
 DR InterPro; IPR003368; -  
 DR Pfam; PF02385; OMP; 1.  
 DR Pfam; PF02415; DUF145; 1.  
 DR Signal.  
 KW Signal.  
 FT CHAIN 17 846 POMP91B.  
 SO SEQUENCE 846 AA; 90834 MW; 4C0DC3IDC0C32964E CRC64;

Query Match 32.1%; Score 1535; DB 2; Length 846;  
 Best Local Similarity 36.4%; Pred. No. 1.3e-82;  
 Matches 350; Conservative 156; Mismatches 306; Indels 150; Gaps 22;

OY 1 MKTSIPWVSSVLAFSCHQ--SLANELLSPDDSFNGNIDSGTFPRK--TSATYSLT 57  
 DB 1 MKHPVWFLLISSSLASNSLSFAOVNTNETLTSSTNGVNTSEEFVKETSGAIITTCBG 60  
 OY 58 DVFYEPGKGTPLSDSCFKOTDNLTLFNGHSLTFEGFIDA-GTHAGAASTYANKLTF 116  
 DB 61 NVCISYAGKDSPLMKSCFSETTENLSFIGNGYLTCEDNITTSQSHGALSVSTKNTLDI 120  
 OY 117 SGPSSL--FDSPTSTVYTTGGCTSLSSAGVNLNIRKLVVAGNFSTADGAIKASFLLT 174

Db 240 SITDFQVIFDGNASAMEAQAQGAICCTTDTK--VLTGNKMLSTNTMTALFYGAIS 297  
 Qy 287 AKKIALSSGGEFTEFLRNNVSSATP---KGAIISIDASGELSASGTGTTVRNLTITG 343  
 Db 298 GLKVISISGGPFLF-QSNISGSSAGGGGAINIASAGELIASATSGDITP--NNQVING 355  
 Qy 344 STDPPKRAINIGSNGKFTELRAKNHTIFFYDPTSEGT--SSDVLKINNGSAGALMPY 401  
 Db 356 STST--RAAINIDTAKTYSIRATGOSIFYDPTINGCTAASDTLNLADANSELEY 413  
 Qy 402 QCTLLFSGEYLTADELKADYADMKSEFTQVPSLSGGKLLQKGVLESTFSQOAGSLIG 461  
 Db 414 GGAIVFSGSEKISPEKAIANAIVTSTIRQPAVLARGDLVLRDGVYTFKDLQSPGSRILM 473  
 Qy 462 DSGTLLSTTASITTTNGLINVDLSGLKQPVSLAKGSMNVYISGKLNLDIEGNIYES 521  
 Db 474 DGGTLLSKKANLSLNLGLAVNLSSLDGTNKAALTEADKNISLSGTLALDITGESFEN 533  
 Qy 522 HMFSDOLFSLKLTVDVDTNVDISLIPVPAEDPSEFGOGQMVNMTTDTATNTK 581  
 Db 534 HNLKASATPYLELT--AGANGTTLGALSTLTLOEPETHGYGQNMOLMA--NATSSK 590  
 Qy 582 EATATWTKTFVSPERKSALVNTLWGVTDIRSLQOLVEIGATGMBHKGQFWVSSMTN 641  
 Db 591 IGSINMTFTGYIIPSEKSNPLUNSLMGNFIIDINSINGLIFETKSGSEPEBELMLSGIAN 650  
 Qy 642 FLHKTGDNKRGFRHTSGGYIGSAHTPKDDLTPFACHLPARDKOCFIANNSTRYGG 701  
 Db 651 FFYDSMPTRHGFRHISGVALGITATTPADDLTFAFCOLFARBRNHTGKNHDTYGA 710  
 Qy 702 TLFRKSHHTL-QOPNYLRLGAKFSESALIEKPREIPALVOVSFSDRMETHYSL 760  
 Db 711 SLYHHTHREGLDIANFL-WGKATRAPVWLSISQIPLSPAKFSYLTHTDHMKTYIDN 769  
 Qy 761 PESGSMNSNECICAGIGIDLPVLSNPHLEKFTIIPOMKVEMVYVSONSFESSDGRGF 820  
 Db 770 SIIGSMWDAFCADLGLSLPFLVISVP--LLKEVEPEKVVYIAHOODFEREYAEGRAF 828  
 Qy 821 SIGRLNLSIPVGAKKFVGODIGDSYTDLSGFYSDVYRNNPOSTATILYMPDMSKIRGG 880  
 Db 829 NKSELINVEITGYTFERDSKSEKGYDILMTYILDAIRNPKCOTSLIADAMMAYGT 888  
 Qy 881 NLSROAFLRGSNNVYVNSNCELFHGMELRGSNNVYVNGTKRLR 928  
 Db 889 NLAQGSVRAANHFQVPHMEIFGQFAFEVRSRRNNTMLSKFCF 936  
 RESULT 11  
 Q9RB71 PRELIMINARY; PRT; 772 AA.  
 AC Q9RB71:  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE PMP.3.  
 GN PMP.3.2.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. *Chlamydia*.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138:  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CML029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL, AF002545; BA98226.1; -.  
 DR InterPro: IPR003357; -.  
 DR InterPro: IPR003368; -.  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.

SQ SEQUENCE 772 AA; 82931 MW; 484FC56D635801EB CRC64;  
 Query Match 33.38; Score 1591; DB 2; Length 772;  
 Best Local Similarity 42.78; Pred. No. 5.7e-86;  
 Matches 337; Conservative 146; Mismatches 275; Indels 32; Gaps 14;  
 Qy 153 LVVAFNESTAGGAIKGSFLLTGTSGDALFNNSSSTKGAATTAGARIANNVGYRF 212  
 Db 1 MLFSKNEFTDNGAIIATKLTSLGTGMSALFSEMTSSKKGAIQTSIDLITTGNGEVSF 60  
 Qy 213 LSNIASTSGAIDEGTSLISNNKFLYF----EGNAKTT----GGAICNTKASGDEL 263  
 Db 61 SDNMSDGAALFEASVYISNNAKVSPIDNKYVAGASSSTGDMSGAICAYKTSITDKV 120  
 Qy 264 IISNKKTLIFASNAVETSGGAIHAKKIALSSGGEFTEFLRNNVSSAT--PKGAISIDASGE 322  
 Db 121 TLTGQMLFLFNNMTSTPAGAIYVKKLELAGGLTFERNVNGTAPKAGAIADDSGE 180  
 Qy 323 LLSAETGNTFFVNTLTGTTSTDPKRNAINIGSNGKFTELRAKNHTIFFYDPT--S 380  
 Db 181 LLSLSDSGDIVFLGNTVYST--TPEGRNSSIDLGSAMKALRSAGRAIIFDPTTGS 238  
 Qy 381 EGTSDVLEKINNGSAGALNPYQGTILFSGETLTADDELKVADNLKSSFTQVPSLSGKLL 440  
 Db 239 STYTVDLKAVNETPADSALOYTGNIIFTGEKLSFTEADSKNLISKLQVTLTSGTSL 298  
 Qy 441 QKGVLTSTFSQOAGSLIGMSGTTSTAGSITTNGLINVDLSGLKQPVSLTAKGAS 500  
 Db 299 KHGYLTQAFVTOQADSLEMDVGTTLT-PADTSTINLVYIISIDGAKKAKLETNAT 357  
 Qy 501 NKVIYSGKLNLDIEGNTYESHMSHDLFSLKITVADVDVYNDISLIPVPAEDPNS 560  
 Db 358 KNLTLSTGTLTLDPTGTFEYENHSLRNPOSYDILEKASGYTS----TAVPDDPMGKRF 413  
 Qy 561 EYFGQGMN-VNMTTDTATNTKEATATWTKTFVSPERKSALVNTLWGVTDIRSLQ 619  
 Db 414 HGYQGTWGPVWGTAASYT--ATFNMTKGYIIPNERIGSLVPSNLMAFIDISSLYH 470  
 Qy 620 LVEIGATGMEKQGFVWSSMTPFLKGTGDNKRGFRHTSGGYIGSAHTPKDDLTPFAF 679  
 Db 471 LMETANEGLQDRAFWCAGLSNPFHKDSTKTRRGFRHLSGGYVIGNLITCSDKILSNAF 530  
 Qy 680 CHLPARDKOCFIANNSTRYGTLEFFKSHHTLOPONYLRLRAKFSALIEKPREIPLA 739  
 Db 531 COLFGRDRDYAAKNQGVYIGTLYQHNET-----YISL-PCILRPSLSYVTEIPVL 584  
 Qy 740 LDVQVFSHSDNRMETHYSLPESGSMNSNECICAGIGIDLPVLSNPHLEKFTIPIOM 799  
 Db 585 FSGNLSTHTNDLKTXYTTPYVKGSMGNSFALFEFGRAPICL-DESALEFQYMPFMK 643  
 Qy 800 VEMVYVSONSFESSSDGRGFSIGRLNLSIPVGAKF-VQGDIDISTYDLSGFVSDVY 858  
 Db 644 LQFYAHQEGFEQGTETAREFGSSRLVNLALPIGIRPEKSDCCDA--TYNLTJLYTYDLV 702  
 Qy 859 RNNPOSTATILVSPDSMKIRBGNLSROAFLRGSNNVYVNSNCELFHGMELRGSNNY 918  
 Db 703 RSNPDCTTLRLISGDSWKTFTGTLNLRQALVLRAGNHCFNSNFEASOFSELMGSSNNY 762  
 Qy 919 NVDVGTFLRF 928  
 Db 763 NVDGAKYQF 772  
 RESULT 12  
 P77792 PRELIMINARY; PRT; 839 AA.  
 AC P77792:  
 DT 01-FEB-1997 (Tremblrel. 02, Created)  
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE POMPOB PRECURSOR.  
 OS Chlamydia psittaci (Chlamydia psittaci).

RL Nat. Genet. 21:385-389(1999).  
 DR EMBL: AE001627; AAD18589.1; -.  
 DR InterPro: IPR003357; -.  
 DR InterPro: IPR003358; -.  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.  
 DR SEQUENCE 936 AA; 100079 MW; 8881D78A53D194EC CRC64;

Query Match 36.3%; Score 1734; DB 2; Length 936;  
 Best Local Similarity 39.9%; Pred. No. 2.7e-94;  
 Matches 378; Conservative 182; Mismatches 356; Indels 32; Gaps 18;

1 MKSIPWLVSVYAFSCHLQSLANEELSPDPSFNGNIDSGT-----TPKTSATYSL 55  
 1 MKSSVSMLEFSSSIPLRSSLSIVAEEVTLDDSSNNSYDS -NGTFYFYSTDAAGTYYSL 59  
 56 TGDVFEYEPK -GTPLSDSCEKOTDNLTFLGNGHSLTFEGFDAGTHAGAASTT-ANKN 113  
 60 LSDVSPQNALGALGPLASGCFLEAGGDLTFQGNQHALKFAPINAGSAGVASTSAADKN 119  
 114 LTFSSGSLSFDSSTPTTYV-TGCGTLSSAGVNLNIRKLVVAGNFTADGAIGASF 172  
 120 LTFNDSRLSIIISCPSSLPTGQCALKSVGNLSLTGNSOIIFTONFSSDNGVIMTKNF 179  
 173 LTFGTSGDALFSNNSSST--KGAIAATTAGARIANNMGVRFSLNSTASTSGAIDDEGTS 230  
 180 LLSGTQSFASFNRQAFTGQGGVYATGTTITENSFGIVSEFQNLAKSGGALYSTDNC 239  
 231 ILSNNKFLYEGN-----AAKTGGAICNTRKASGSPELIISNNKTLIFASVNAETSGAII 286  
 240 SITDNOVIFDGSNMAEAQAQACCTTDTKT--VTLGKNNLSPTNNLTALYGAIS 297  
 287 AKLALSSGGEFTEFLRNNSATP---KGAISIDASGELSLSAEGNITTFVNTLTGTTG 343  
 298 GLVYSISAGPTLF-OSNISGSAGOGGGAIIASAGELATLSQDITF-NNQVTTNG 355  
 344 STDPKRNAININISNGFTELRANKNHTTFEYPTISEGT--SSDYKINNGSAGALNPY 401  
 356 STST--RNAINITIDTAKVISIRATGOSITFYPTINPGIASTDTLNLMLANSSTIEY 413  
 402 QGTTLFSGETLTADELKVDNLKSFQPVYLSGSKLLQKCVTLSTSPSOBAGSLMG 461  
 414 GGAIVSGEKLSPTEKAIANAVTSTIRQPAVALRGDLVLDGTVTPKDLTQSPRIIM 473  
 462 DSGTTLSTGTAGSTITNLGINVDSLAKQPVSLTAKGASKVYSGKLLIDEGNIYES 521  
 474 DGGTTLTSAKEANLSLNGLVANLSLQGTNKAALKTEADNNISLSGTIALIDTEGSEYEN 533  
 522 HMFSDQLFSLKITYDADVDYVNDISLIPVADPNSEYFGQGNVMMWTDTATNTRK 581  
 534 HNKASASTYLLLELT-AGANGTITLALSTLLOEPETHYGOGNQOLSLMA--NATSSK 590  
 582 EAAATWTKTGFVSPERKSLAVCNTLMGVFTDIRSLQOLVEIGATGMEHKQGFVSSMTN 641  
 591 IGSINWTRCYIYSPERKSNMLPLNSLMGNFIDIRISNQLLETFSSGPFERREMLSGIAN 650  
 642 FLKKTGDENKGGRRHSGGVIGVSGAHTPEKDLFEAFCHLFARDKCFIAHNNSRTYGG 701  
 651 FFFRDSMPTRHGRHISGGVAGLDTATTPAEDQLTFACDLFARDNRHITGKHNGDYGA 710  
 702 TLFKKHSHTL-QPQNYLRGLRAKFSASALEKFPREIPLADLVQVSSHSNDRMETHYSL 760  
 711 SLVFNHTEGLFIANLF-WKATRAPWVLTSEISQIILPSDAKFSYLHNDNHKITYTLDN 769  
 761 PSEBGSNSNCIAGIGLDPYLSNPHLLEKTFIPOMKEMVYVSQNSSESSDGRGF 820  
 770 SIKGSNRNDFACADIGASLPFYISVPY-LKLEVEBPVKKQYIYADQDYEERHAGRAE 828  
 821 SIGRLNLTSIPVAKKFGQDIDSYTYDLSGFFVSVYRNPNPOSTATLYWSPDSMKIRGG 880  
 829 NKSELIWELIPGVTERDSKSEKGYTDLTMTLIDVYRNPRCQSTLSLASDANMAAYGT 888

OY 881 NLSROAFLEGRSNNVYNSNCELFCHYAMELRGSSRNYNVDGTKLRF 928  
 889 NLRQGFSVAAHHFOVNPIMELFGQFAFEYVRSSSRNYNNILGSKICF 936

RESULT 10  
 09JS42  
 ID 09JS42 PRELIMINARY; PRT; 936 AA.  
 AC 09JS42;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, last annotation update)  
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE  
 DE PROPEIN G FAMILY).  
 GN PMR\_7 OR CP0308.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138.  
 RC MEDLINE=20330349; PubMed=10871362;  
 RA Shital M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWI029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwin M., Nelson W., DeBoy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MOpn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AF002546; BAA98653.1; -.  
 DR EMBL: AE002193; AAF38165.1; -.  
 DR TIGR: CP0308; -.  
 DR InterPro: IPR002637; -.  
 DR InterPro: IPR003357; -.  
 DR InterPro: IPR003358; -.  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.  
 DR ProDom: PD004952; -; 1.  
 DR SEQUENCE 936 AA; 100105 MW; 3981DB3C950AF95A CRC64;

Query Match 36.2%; Score 1733; DB 2; Length 936;  
 Best Local Similarity 39.9%; Pred. No. 3.1e-94;  
 Matches 378; Conservative 182; Mismatches 356; Indels 32; Gaps 18;

1 MKSIPWLVSVYAFSCHLQSLANEELSPDPSFNGNIDSGT-----TPKTSATYSL 55  
 1 MKSSVSMLEFSSSIPLRSSLSIVAEEVTLDDSSNNSYDS -NGTFYFYSTDAAGTYYSL 59  
 56 TGDVFEYEPK -GTPLSDSCEKOTDNLTFLGNGHSLTFEGFDAGTHAGAASTT-ANKN 113  
 60 LSDVSPQNALGALGPLASGCFLEAGGDLTFQGNQHALKFAPINAGSAGVASTSAADKN 119  
 114 LTFSSGSLSFDSSTPTTYV-TGCGTLSSAGVNLNIRKLVVAGNFTADGAIGASF 172  
 120 LTFNDSRLSIIISCPSSLPTGQCALKSVGNLSLTGNSOIIFTONFSSDNGVIMTKNF 179  
 173 LTFGTSGDALFSNNSSST--KGAIAATTAGARIANNMGVRFSLNSTASTSGAIDDEGTS 230  
 180 LLSGTQSFASFNRQAFTGQGGVYATGTTITENSFGIVSEFQNLAKSGGALYSTDNC 239  
 231 ILSNNKFLYEGN-----AAKTGGAICNTRKASGSPELIISNNKTLIFASVNAETSGAII 286

```

Db 537 -NOAFTQPLVVFATAAADIYIDALLTSPVQTPPEPHYQGQWEATW-ADTST-AKSGT 593
QY 585 ATWTKGVVSPSEKRSALVCNTLMGVFDIRSLQOLVETGAGMHKQGFVWSSMTNPLH 644
Db 594 MTWVTGTGNPNDEKRSASVPDSLMASTFDIRLQOIMTSQASIIYQOQGLWASGTANPEH 653
QY 645 KTGDENRKGFRTSGGIYIGGSAHFPKDDLFEEAFCHLFARDQCFIAHNNRTYGGTLE 704
Db 654 KDKSGTNGAFRHKSGYIVGSGAEDFSENIPVAFQCLFGKDKDLFIYENTSHNTLASIX 713
QY 705 FRHSTLOPONTLRGKRAKFSESAIEKFPREIPLADVQVSESHDNRMETHYSLPSE 764
Db 714 LQHRFAFLG-----GLPMPSEGSTIDMLKDIPILINAOLSYSTKKNMDRTYTSPEAO 766
QY 765 GWSMECTIAGIGLDLPVLSNPHLPKFTPIPMKVEWVYVYSONSEFPSSSDGREGFSTGR 824
Db 767 GSWTNSGALIEGGLALYLKREAPFGGYRPEPLFQAVYSHQONFKSGAARAFDDOD 826
QY 825 LINTSIPVAKFVQGDIDDSYTYDLSGFFVSDVYRNNPOSTATLVMSPDWKRINGNLNR 884
Db 827 LVNCSIPVIRLEKISEDEKNFETSLAYIGDYVRKNRPSRSLWASGASWTSGLCKNLAR 886
QY 885 QAFILRGSNNVYVNSNCLFGHYAMELGRSSRNNAVDTKILRF 928
Db 887 QAFILASGSHLTLSPHVELSGEAAVELRGSAAHYVNDGCLRYSF 930

RESULT 8
086163 PRELIMINARY; PRT; 914 AA.
ID 086163
AC 086163;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE OUTER MEMBRANE PROTEIN 5 PRECURSOR (FRAGMENT).
GN OMP5.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD/CWL-029/VR-1310;
RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RT "Identification of two novel genes encoding outer membrane complex
associated surface layer proteins in Chlamydia pneumoniae.";
RL Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ001311; CAA04671.1; -.
DR InterPro; IPR003357; -.
DR InterPro; IPR003368; -.
DR Pfam; PF02385; OMP; 1.
DR Pfam; PF02415; DUF145; 1.
KW Signal.
FT SIGNAL 1 15
FT CHAIN 16 >914 POTENTIAL.
FT NON_TER 914 OUTER MEMBRANE PROTEIN 5.
SQ SEQUENCE 914 AA; 95603 MW; 8BF33BAB680FF5E3 CRC64;

Query Match 36.4%; Score 1739; DB 2; Length 914;
Best Local Similarity 40.0%; Pred. No. 1.3e-94;
Matches 376; Conservative 171; Mismatches 341; Indels 52; Gaps 18;

QY 1 MKTSIPWLVSVLA--FSCHLQSLANELLSPPDSFNGNDISGFRTKTSAT--TYSLT 56
Db 1 MKSQFSWLVLSYTLACFPSCSTVFPAATENIGPSDSFGSTNGTGTYYTKNTTIGDYLTF 60
QY 57 GDVFEYEPGKGTPLSDSCFKOTDNLTLFLNGHSLTFEGFDAGTHAGAASATYANKNLTF 116
Db 61 GDITLQNGDRAALTKGCFDSTBESLFRAGKYSLSFLINSSAR-GAALSTYTDKNLSL 119
QY 117 SGFSLSDSSPSTVTTT--GCGTILSAGVNLNIRKLVAAGNFSTADGAIGKASFL 174
Db 120 TGFSSILFLAAPSIVTTPPSGKAVKCGDILFDNNGTILFKQDYCEBNGCAISIKNLSL 179

```

```

QY 175 TGTSGDALFSNNSST---KGAIAITAGARIANTGTGVRFSLNSTASGAIIDEGTSI 231
Db 180 KNSTSISIEBGNKSSATGKKGALCATGTVDTTNTATPLFSNNIAEAAGAINSTGCKT 239
QY 232 LSNNKFLYFEGNAAKT---GGAICNTRASGPELLISNNKTLFASNAVETSGAIIAK 288
Db 240 ITGNTSLFSENSVYATANGCAL-----SGDADVTISGNSVTFSGAQAIVANGAIYAK 294
QY 269 KIALSS--GGTFEPLRNNVSSATP-KGAISIDAGELSLSAETNITFVRRITLTST 345
Db 295 KLTLASGGGGGIFSNNIVQGTAGNGAISILAECSLSAEADITFENGVAIVAT-TP 353
QY 346 DTPKRNAINIGSNGKFTELRAAKNHTIFFYDPTSE--GTSDDYKINNGSAGALNPYG 403
Db 354 QTKRNSIDIGSTAKITNLRASISHSIFFYDITANTADSDITLNLKADAGNSTDVS 413
QY 404 TILFSGETLTADELKVNALSSFTQVPSLSGGKLLLOKVLTSESTFSQEAAGLLGDS 463
Db 414 SIVFSGEKLSDEAKVADNLSTLKQVTLTAGNLVLRKGVTLDRKGFOTAGASSVIDA 473
QY 464 GTTLESTTAGSITITNLGINVDSLGKQVSLAKASKKYVSGKLNLDIEGNIYESHM 523
Db 474 GTTLKASTEETLTGLSIPVDSLGKKRVLAASAASKNVALSGPILLLDQGNAYENHD 533
QY 524 FSHDQFLSLKITYVDADVDTNVDISLIPVPAEDPNSEYFGQGNVWMTDTAT--NTK 581
Db 534 LKTDQDFEFVQLSA-LGATITTDVPA---VPVATPTHTGYGTGMWVDDTASTPTK 589
QY 582 EATATWTKGVVSPSEKRSALVCNTLMGVFDIRSLQOLVETGAGMHKQGFVWSSMTN 641
Db 590 TATLATWNTGYLPNDEKRGPIVPSNLGCSFSDIAQIGVIERSAITLDCSDGFMAAGVAN 649
QY 642 FLHKRGDENRKGFRRTSGGIYIGGSAHFPKDDLFEEAFCHLFARDQCFIAHNNRTYGG 701
Db 650 FLDKRKGKRRYRHRSGGYALGGAQTCSENLISFAFCQLFGSKDPLVAKNHTDIAG 709
QY 702 TLFFKSHSTLQPONTLRGKRAKFSESAIEKFP--REIPLADVQVSESHDNRMETHYT 758
Db 710 AFYIOHTEC-----SGFICGLDCLKLPGSMKHPVLEGLAASHVNDLKTXYT 759
QY 759 SLPESEGSWSNECIAGIGLDLPVLSNPHP---LKFETPIPMKVEWVYVYSONSEFPSS 814
Db 760 AYPEKFSGSMGNAAFMMAGAS---SHSYPEYLHCFTYVAPYIKLNTLYIROSFSSEK 814
QY 815 SDGRFSISGRLLNSTIPVAKFVQGDIDDSYTYDLSGFFVSDVYRNNPOSTATLVMSDS 874
Db 815 TEGRSFDDSNLFNLSLPIGVKFEKSDCNDPSYDILTISYVDPDLIRNDKCTTALVISGAS 874
QY 875 WKIRGNLSROAFLLRGSNNVYVNSNCLFGHYAMELRGS 914
Db 875 WETVANNLARQALQVRAGSHYAFSPMEVLOQFVEVAGS 914

RESULT 9
092898 PRELIMINARY; PRT; 936 AA.
ID 092898
AC 092898;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE POLYOMORPHIC OUTER MEMBRANE PROTEIN G FAMILY.
GN PMP-7.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";

```



```

Db 1 MKIPLHLILSLTLPPI-LLSIATYCADASLSPDSDG-AGGSTFTPKSTADANTNY 58
QY 54 SLTGDVEFFEPGKGTPLSDSCFKQTDNLTFLNGHSLTFGEFDAGHAGAASSTANKN 113
Db 59 VLSGNVYINDAGGATLTCGCTETTGDTLFTGKGYSFSPNTYDAGSNAGAASSTADKA 118
QY 114 LTFSGFSLSPDSSPTVTYTGQTLSSAGGVLENIRKLVVAGNESTA---DGAATKA 170
Db 119 LTTTGSNLSFIAPGTVAASGSLSSAGALNLTGDTLLFSGONVSNENNNNGATTK 178
QY 171 SFLTGTSGDALFSSNNSSSTKGALATTAGARIANTGYVRLSNIASTSGAIDEGTS 230
Db 179 TLSISGNTSITFTFSNAGKIGAIYSSAAASISGNTGQLVFMNNGETGGALGFEASS 238
QY 231 ILSNNKFLYEGNAKTT---GGAICNTKASGSPELIISNNKTLIFASNAEISGGAIIHA 267
Db 239 STGNSSLFSSGNTATDAAGKGAICYCKTGTEPTTLISGNKSLTFAENSSVTGGAIICA 298
QY 288 KKLALSSGGEFTFLRNIV-SSATPKGGAISIDASGELSLSAETGNTFTVFNLTGTTGSTD 346
Db 299 HGDLISAAGTFLSSNNCGTAAAGKGAIALADSGSLSSANOGDITFLGNTLTST-SAP 357
QY 347 TPKRNAINIGSNCKFTELRAAKNHTTFEYDPTISEGT-SSDVLKINNGSAGALNPYOGTI 405
Db 358 TSPRNAIYLSSAKITFLRAAGOSIIFYDPIASNTGTAGADVLTINQPDNSPLDYSGTI 417
QY 406 LPSGFTLADLKVADNLKSFQPVSLSGKLLKQGVLTLESTFSQEGSLIGMDSGT 465
Db 418 VFSGEKLSADEAKAADFTSILKOPPLASGTLAKGNVELDVNGFTQTEGSTLLMQPGT 477
QY 466 TLTSTAGSITITNGINVDLSLGLKOPVSLTAKGASNKVIVSGKINLIDIGNTYESHMF 525
Db 478 KLRADTAISITLKVYDLSALBGNKSVISITAGANKITITLSPLVFDDSGNFTESHIT- 536
QY 526 HDOLFSL-LKITYDADVTNVDISSLIPVAEDPNSEYFGOGMNVNMTDTATNTKEAT 564
Db 537 -NQAFTQPLVVFATPAATMAADYIDALLTSPVQTPPEPHYGQHEATW-ADTST-AKSGT 593
QY 585 ATMTKTFEVSPEKSKALVONTLMGVTDIRISLQQLVEIGATGHEHOGVWVSMTFLH 644
Db 594 MTWTTTGYNPNPRASVVDLSMASTDIRTQOIMTSGANSISTYQOGRMASTGAFHF 653
QY 645 KTGDENKRGFRHTSGGVYIGSATHPKDLETFEACHLFAADKCFIAHNSRFTYGGTLF 704
Db 654 KDSGTQNAARHRSYGVYIGSADFESNIFSAFOQLPKKDLFIYEMTSHNYLASLY 713
QY 705 FKSHHTLOPNTLRLGAKFSESAIEKFPREIPLADVOYFSHSDNRMTHTYSLPESE 764
Db 714 LQHRALFG-----GLMPSPFSITDMLMDIPLILNAOLSYSTKNDMDTRYSYEAQ 766
QY 765 GSNMNECIAGIGIDLPFVLSNPHLPKFTIPOMKVENYVVSNSPSSSDGSGFSGR 824
Db 767 GSWTNSGALLEGSLALYLPKPAEPFOGYFPPLKFOAVYSROONFESASAEAAAFDGD 826
QY 825 LNLNLSFVGAKFYOGDIGDYTDLSGFEVSDYVRNPNPOSTATILVPSDMKIRGNMLSR 884
Db 827 LVNCSIVVGRLEKISDEKNNFEISLAYIGDYVRKPNRRTSILMVGASAWTSLCKLAR 886
QY 885 QAFLLKGSNNYVNSNCELFGHYAMELRGSSRNIVDVGYTKLR 928
Db 887 QAFLASAGSHLTLSLPHVELSGEAAVELRGAHYIYNDGCHRYSF 930

```

```

OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria: Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
OX NCBI_TaxID=83558;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-CML029;
RX MEDLINE=9206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lamell C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.",
RL Nat. Genet. 21:385-389(1999).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN-VRI310;
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birkeland S.;
RT "Chlamydia proteins containing the GGAI-repeat belong to a subfamily
RT of autoantigenic pathogenicity factors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001627; AAD18590.1; -
DR EMBL; AJ133034; CAB37068.1; -
DR InterPro; IPR003357; -
DR InterPro; IPR003358; -
DR Pfam; PF02385; OMP; 1.
DR Pfam; PF02415; DUF145; 1.
KW Signal.
FT SIGNAL 1 27
FT CHAIN 28 930
FT SEQUENCE 930 AA; 97669 MW; 46A9B5E3BB913C4C CRC64;
SQ

```

Query Match 36.8%; Score 1761; DB 2; Length 930;  
 Best Local Similarity 41.7%; Pred. No. 6; 9e-96;  
 Matches 394; Conservative 165; Mismatches 355; Indels 30; Gaps 14;

```

QY 1 MKTSLPWLVSVALESLQSLAN---EELLSPDSDSGNIDSGTFTPTSA---TTY 53
Db 1 MKIPLHLILSLTLPPI-LLSIATYCADASLSPDSDG-AGGSTFTPKSTADANTNY 58
QY 54 SLTGDVEFFEPGKGTPLSDSCFKQTDNLTFLNGHSLTFGEFDAGHAGAASSTANKN 113
Db 59 VLSGNVYINDAGGATLTCGCTETTGDTLFTGKGYSFSPNTYDAGSNAGAASSTADKA 118
QY 114 LTFSGFSLSPDSSPTVTYTGQTLSSAGGVLENIRKLVVAGNESTA---DGAATKA 170
Db 119 LTTTGSNLSFIAPGTVAASGSLSSAGALNLTGDTLLFSGONVSNENNNNGATTK 178
QY 171 SFLTGTSGDALFSSNNSSSTKGALATTAGARIANTGYVRLSNIASTSGAIDEGTS 230
Db 179 TLSISGNTSITFTFSNAGKIGAIYSSAAASISGNTGQLVFMNNGETGGALGFEASS 238
QY 231 ILSNNKFLYEGNAKTT---GGAICNTKASGSPELIISNNKTLIFASNAEISGGAIIHA 267
Db 239 STGNSSLFSSGNTATDAAGKGAICYCKTGTEPTTLISGNKSLTFAENSSVTGGAIICA 298
QY 288 KKLALSSGGEFTFLRNIV-SSATPKGGAISIDASGELSLSAETGNTFTVFNLTGTTGSTD 346
Db 299 HGDLISAAGTFLSSNNCGTAAAGKGAIALADSGSLSSANOGDITFLGNTLTST-SAP 357
QY 347 TPKRNAINIGSNCKFTELRAAKNHTTFEYDPTISEGT-SSDVLKINNGSAGALNPYOGTI 405
Db 358 TSPRNAIYLSSAKITFLRAAGOSIIFYDPIASNTGTAGADVLTINQPDNSPLDYSGTI 417
QY 406 LPSGFTLADLKVADNLKSFQPVSLSGKLLKQGVLTLESTFSQEGSLIGMDSGT 465
Db 418 VFSGEKLSADEAKAADFTSILKOPPLASGTLAKGNVELDVNGFTQTEGSTLLMQPGT 477
QY 466 TLTSTAGSITITNGINVDLSLGLKOPVSLTAKGASNKVIVSGKINLIDIGNTYESHMF 525
Db 478 KLRADTAISITLKVYDLSALBGNKSVISITAGANKITITLSPLVFDDSGNFTESHIT- 536
QY 526 HDOLFSL-LKITYDADVTNVDISSLIPVAEDPNSEYFGOGMNVNMTDTATNTKEAT 564

```

RC STRAIN-VRI310;  
 RA Hierno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,  
 RA Christiansen G., Birkelund S.;  
 RT "Chlamydia proteins containing the GGAI-repeat belong to a subfamily  
 RT of autoantigenic pathogenicity factors";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umeyam L.A., Utterback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwinn M., Nelson W., Debroy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AP002546; BAA98657.1; -;  
 DR EMBL: AJ133034; CAB37071.1; -;  
 DR EMBL: AE002192; AAF38160.1; -;  
 DR TIGR: CP0303; -;  
 DR InterPro: IPR003357; -;  
 DR Pfam: PF02385; OMP; 1.  
 DR Signal.  
 SO SEQUENCE 928 AA; 97229 MW; 0590D5206A1DD0E1 CRC64;

Query Match 37.38; Score 1785; DB 2; Length 928;  
 Best Local Similarity 40.48; Pred. No. 2.6e-97;  
 Matches 385; Conservative 173; Mismatches 344; Indels 52; Gaps 18;

1 MKTSPWLVSVLA--FSCHLQSLANELLSPDSFNGNIDSGFTPKTSAT--TYSLT 56  
 1 MKSPQMLVLSSTLACFCSCSTVPAATAENIGPDSFSGSNTGTYTPKNTTGIDITLT 60  
 57 GDVEFFEPGKPLSDSCFKQTDNLTLFLNGHSLTEGFDAGTHAGAASTANKLTF 116  
 61 GDITLQNGDSALATKQGFSDTTELSFPAKGYSLSFLNIMSSAE--GALLVTTDKNLSL 119  
 117 SGFSLSPDSSPTTQVT--GQGLTSSAGVNLNIRKLVYAGNFSTADGAIKAGASFL 174  
 120 TGFSLFLFLAPSSVITTPSGKGVKCGDITLTPNNGITLTKODYCEHNGAISTKNLSL 179  
 175 TGTSGDALFSNNSSST---KGAATATTAGARIANTGTVRPLSNASTSGAIDEGISI 231  
 180 KNSGTGSLSEFGNKSAGTKGGAICATGTVDITNTATPLFSNNIAEAGAINSTGCT 239  
 232 LSNKRFYFEGNNAKT---GGAICNTKASGSPELIISNNKTLIFASVNAETSGAIAHAK 288  
 240 ITGNTSLVFSNSVTATNAGNGAL----SGDADVTISGNSGVTFSGQNAVANGAIYAK 294  
 289 KLAIS--GCFTEPLRNWVSATP--KGAISIDASGELISAEFGNITFVNLTITGSGT 345  
 295 KLTLASGGGGGSISSNNIVQGTAGNGAISILAAEGCSLAEAGDITFNNAIYAT--TP 353  
 346 DTPKRNAINIGSNKFTETLRANKHTTFEYDPISE--GTSDDLKINSGAALNPYOG 403  
 354 QTKRNSIDIGSTAKITNLRLAISGHSIFEDYPIANTRAADSTPLNLKADKADNSTDSG 413  
 404 TILFSGETLPADELKVAADNLKSFOTOPVSLSGKILLQKGVLTSTSPQASGLKADS 463  
 414 SIYVSGKLSLDEKAVNDNLSTLKQPYTLTLAGNLVLRGVTLDTKGTQYAGSSVIMDA 473  
 464 GTTILSTAGSITITNLGINVDSLKQPVSLTAKGASNKVYLSKLNLIIDIEGNIYESHM 523  
 474 GTTLKASTEEVTLTGLSLIPVSLGEGKKVYIASAASKNVALSGPILLDNOGNAEYHD 533  
 524 FSHQULSLKITYTADAVDVTWVDSLIPVPAEDPNSEYGGQGMVNWTTDTAT--NTK 581  
 534 LKQTQDSFVOLA--LGTATTTDPA--VPTVATPHYGQGTWGTWVDSTSTPKTK 589  
 582 EATATWTKTGKVPSPERKSALVCTINGVFTDIDSLQDLVIGATGMEHKGGFVWSKTN 641

Db 590 TATLAWNTGYLPNERPGPLVPNSIMGSFSDIOAQIERSALITLCSDBGFNAAGVAN 649  
 Qy 642 FLHKTGDNKRGCFHTSSGYVIGSAHPKDDLTFFACHIFAPADKCOFTHNNSRTG 701  
 Db 650 FLDDKCKEKKRYHRKSSGYAIGAAQCSENLISFACQGLSGDKDLVAKNHTDYAG 709  
 Qy 702 TLFFKSHHTLQPNVTLRLGRAKFSESAIEKPP---REIPLADVOVSFSHSDNMEHTY 758  
 Db 710 AFYQHTIEC-----SGFTICLLDKLPKSMKHPVLLEQALVSHVSNLKTFTY 759  
 Qy 759 SLPSEGSMSNECIAGIGIDLPPVLSNPH--LEKFTIPQMKVENYVSQNSFESS 814  
 Db 760 AYPEVKGSGWGNNAFMMMGAS-----SHSYPEYLHCFTDYATPAIKNLFTYIROPDSSEK 814  
 Qy 815 SDGRGFSIGRLNLSIPGAKFVQGDIDSTTYDLSGFSVDYVRNNQSTATLVMSDS 874  
 Db 815 TEGRSFDDSNLFNLSLPVGVKEFRKSDCNDPSYDLTISYVDLIRNDKCTTALVISGAS 874  
 Qy 875 WKIRGNLSROAFLLRGSNNYVNSNCELPGHYAMELRLGSSRNTNVDTKLR 928  
 Db 875 WETANNIARALQVLRAGSHAFSPMEVVLGQFVFEVGRSSRIITVLDGKRFQF 928

RESULT 6  
 ID Q9RB66 PRELIMINARY; PRT; 930 AA.  
 AC Q9RB66;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE  
 DE PROTEIN G FAMILY).  
 GN PWP\_8 OR CP0307.  
 OS Chlamydia pneumoniae (Chlamydiophila pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.  
 OX NCBI\_Taxid=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CML029 from USA";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umeyam L.A., Utterback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwinn M., Nelson W., Debroy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AP002546; BAA98654.1; -;  
 DR EMBL: AE002193; AAF38164.1; -;  
 DR TIGR: CP0307; -;  
 DR InterPro: IPR003357; -;  
 DR InterPro: IPR003368; -;  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.  
 SO SEQUENCE 930 AA; 97639 MW; 46A8896761391C09 CRC64;

Query Match 36.98; Score 1763; DB 2; Length 930;  
 Best Local Similarity 41.78; Pred. No. 5.2e-96;  
 Matches 394; Conservative 165; Mismatches 355; Indels 30; Gaps 14;

1 MKTSPWLVSVLA--FSCHLQSLANELLSPDSFNGNIDSGFTPKTSAT--TYSLT 53

OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CW1029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VRL10;  
 RA Hjerto K., Boesen T., Daugaard L., Knudsen K., Madsen A.,  
 RA Christiansen G., Birke Lund S.;  
 RT "Chlamydia proteins containing the GGAI-repeat belong to a subfamily  
 RT of autoantigenic pathogenicity factors";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CW1029 from USA";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Ullrich T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwinn M., Nelson M., DeBoy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 RT pneumoniae AR39";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AE001628; AAD18591.1;  
 DR EMBL: AJ133034; CAB37069.1;  
 DR EMBL: AF002546; BAA96655.1;  
 DR EMBL: AE002192; AAF38163.1;  
 DR TIGR: CP0306; AAF38163.1;  
 DR InterPro: IPR003357;  
 DR InterPro: IPR003368;  
 DR Pfam: PF02385; OMP.1;  
 DR Pfam: PF02415; DUF145; 1.  
 KM Signal.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 928 OUTER MEMBRANE PROTEIN.  
 SQ SEQUENCE 928 AA; 98332 MW; 58910A8F04F12219 CRC64;

Query Match 37.5%; Score 1794; DB 2; Length 928;  
 Best Local Similarity 42.7%; Pred. No. 7.7e-98;  
 Matches 404; Conservative 152; Mismatches 353; Indels 38; Gaps 17;

OY 1 MKSIPVAVSVLAFSCHIQ---SLANELSPDSEFGNIDSGTFNP-----KTSAT 51  
 DB 1 MKSLHMFLLSSSLALPLSLNFSAFAVYINIGPTNSBGG---PTGYTPPAOTTNADGT 57  
 OY 52 TYSLVGDVFEYERKGTPLSDSCFKQTTDNLTFELGNHSLTFGFLDAGTAGAASSTAN 111  
 DB 58 IYMLTGDSVITNAGSPFALJASCFKFTETGNLSFGHGHYQFLQNDAGANC-TFTWTAAN 116  
 OY 112 KNITFSGFSLSDSSSTVTYTGOGTLSSAGVNLNIRKLVYANFESADGAIKAS 171  
 DB 117 KILSFSGFSLSL--IQTNATTTGTAINSTGACISQNSCYFGNFSNDNGALOGSS 174  
 OY 172 FLITGTSGDLFFSNSSSTGATATAGARIANNNTGYVFLSNISTASGAIIDEGTST 231  
 DB 175 ISLS-LNPNTLTFPKKATKGALYSTGTITINNTLSASFSENTAANGAIYTPASSP 233

OY 232 LSNNKFLYFEGN--AAKTGGAI-CNTKASGSPELLISNNKTLIRASVNAETSGAIIHA 287  
 DB 234 ISSNKALISFINSVTAATSAAGCAIYCSTSAAPKPVLTLSDNGLNLTGNATIRSGAIYT 293  
 OY 288 KIALSSGCTEFLRN-VSAPFKGCAISIDASGELSLSAENGTTFVANTLT--TGST 345  
 DB 294 DNVLSSGCTFLKNNASADITPAALGALAIADSGISLSALGDDITFEENTVVKGASSS 353  
 OY 346 DFKRNAINIG-SNGKFTLEAKNHTIFPDITSQGT--SDVLRKNGSAGALNPYQ 402  
 DB 354 QTTTRNSINIGNTAKIVQURASOGNTIHYDITITSITRALSADALNCPDLAAGNPAYQ 413  
 OY 403 GTILFSGETTLADELVADNIMKSSFTQVSLSGCKLLQKGVTLLESTFSQEAQSLGMD 462  
 DB 414 GTILFSGEKSEAEAEADMLKSTIQGPLLADGQLSLKGVTLVAKSFQSGPSTLMD 473  
 OY 463 SGTTLSTTAGSITTYNLGINVDSLGLKOPVSLAKGASKNVYISKINLIDISGNIYESH 522  
 DB 474 AGTTLLEADG-ITINMLVNVDSLKETKATLKATQASQTVTLTSGISLVDPSGNYEDV 532  
 OY 523 MESHDLFSLKITVDADVDTNVDISLIPVPAEDPSEYFGQGMNMTTDTATPTXE 582  
 DB 533 SWNNPOVFSLTFL--ADDPANIHITDLAADPLEKNPIMHGYGNMALSQEDPTATKSKA 590  
 OY 583 ATATWTKTGFVPSPERKSALVCLNGVFTDIRSLQOLVEIGATGHEHKGQFVWSSMTNF 642  
 DB 591 ATLWTWKTGYNPNRPERGLIVANTMGSEFVDRSIOQLVATKVRQSGETGICEGISNF 650  
 OY 643 LKFTGDNKRGFHTSGGYITGSAHTPKDDLTFAFCHLFAADKCFIAHNSRTYIGT 702  
 DB 651 FHKDSIRKINGFHISAGVYVAGTTLASDNLITTAFCQLFGDRDHFINKNNAASYAAS 710  
 OY 703 LFKHSTLLOPOVYLRGRKAFSESAIEKPPREIPALDVQVFSNDNMEHTYSLPE 762  
 DB 711 LHLQHLATLSPBLK--LPGSES-----EQVLFDAQIYIYSKNTMKRYYYQAPK 761  
 OY 763 SEGSWSNECTAGIGLDLPFVLNPHLEKFTIPQMKVENVYQSNSFEFSSD-GRFS 821  
 DB 762 GESSWYNDGCALCLASLPFALSHGCLFPAVFPFKIVEASYIHODSFKRNTTLVRSFD 821  
 OY 822 IGRLLNLSIPVGAKFVQGDIGDSTYDLSGFEVSDVYRNNQSTATLYPMSPMKIRGN 881  
 DB 822 SGDLINVSPIGITFEFSSNERNASTYATYIYADVYRKNPDCDTTALLNNTSKMTTGIN 881  
 OY 882 LSRQAFLLKGSNNYVYNSNCELFGHAMELKSSRNVDVGTKLRF 928  
 DB 882 LSRQAGIGRAGITFATSPNLEVTNSLMEIRGSSRSRYNADLGCKPQF 928  
 RESULT 5  
 ID 09RB65 PRELIMINARY; PRT: 928 AA.  
 AC 09RB65; 09RB64; 09S6P2; 13, Created)  
 DT 01-MAY-2000 (TRENBLREL, 15, last sequence update)  
 DT 01-OCT-2000 (TRENBLREL, 16, last annotation update)  
 DE POLYMERASE OUTERMEMBRANE PROTEIN G FAMILY (OUTER MEMBRANE PROTEIN 5  
 DE PRECURSOR) (POLYMERASE MEMBRANE PROTEIN G FAMILY).  
 GN PMP\_10 OR OMP5 OR CP0303  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.  
 ON NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CW1029 from USA";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.

OY	361	FTELPAANHIIIEFDPIETSGTSSDYKINNGSAGLNPVOCILFISGFTIADDELKYA	420
Db	362	FTELEAAHNHTIEFDPIITSGTSSDYKLINGSGALNPNPQGITLFSGFTIADDELKYA	441
OY	421	DNLKSSFOPVPSLSGGKLLDQKVTLLESTSPSOEAGSLIGMDSGTVLTSTAGSTITTNLG	480
Db	442	DNLKSSFOPVPSLSGGKLLDQKVTLLESTSPSOEAGSLIGMDSGTVLTSTAGSTITTNLG	501
OY	481	INVDSLGIKOPVSLTAGKASKNVIVSGKLNLIIDEGNIYESHMFSDQLFLSKITVDAD	540
Db	502	INVDSLGIKOPVSLTAGKASKNVIVSGKLNLIIDEGNIYESHMFSDQLFLSKITVDAD	561
OY	541	VDTNVDISLIPVAPEADNSXYGOGOMNNWTTDTAFNTKEATATWTKTGFVPSPERKS	600
Db	562	VDTNVDISLIPVAPEADNSXYGOGOMNNWTTDTAFNTKEATATWTKTGFVPSPERKS	621
OY	601	ALVCNTLWGVTFTDIRSLDQQLVEIATGEHKOGFWWSMNFLLKHTGDENRKGFRHTSGG	660
Db	622	ALVCNTLWGVTFTDIRSLDQQLVEIATGEHKOGFWWSMNFLLKHTGDENRKGFRHTSGG	681
OY	661	YVIGGSATHPKDDLTEFAECHLEARDKDCFIAHNNSRTYGGLTFEKHSHTLOPNYLRLG	720
Db	682	YVIGGSATHPKDDLTEFAECHLEARDKDCFIAHNNSRTYGGLTFEKHSHTLOPNYLRLG	741
OY	721	RAKSESSEIAEKFPREIRPLADVVSFSHSDRMETHYTSLPESGSGSMNCIAGGIGLDL	780
Db	742	RAKSESSEIAEKFPREIRPLADVVSFSHSDRMETHYTSLPESGSGSMNCIAGGIGLDL	801
OY	781	PFLVLSNPAPLPKETEIPQMKVMYVVSONSFESSSDGRGFSIGRLNLSPVGAKFYQGD	840
Db	802	PFLVLSNPAPLPKETEIPQMKVMYVVSONSFESSSDGRGFSIGRLNLSPVGAKFYQGD	861
OY	841	IGDSYTYDLSCGFEPVSDVYRRNPPOSTATLVMSPDSDKIRGNLSROAEFLRGSNYYNSN	900
Db	862	IGDSYTYDLSCGFEPVSDVYRRNPPOSTATLVMSPDSDKIRGNLSROAEFLRGSNYYNSN	921
OY	901	CELRGHYAMELRGSSRNRYNDVGTCKLRF	928
Db	922	CELRGHYAMELRGSSRNRYNDVGTCKLRF	949
<hr/>			
RESULT 3 .			
P71135 PRELIMINARY; PRT: 926 AA.			
ID	P71135		
AC	P71135;		
DT	01-FEB-1997 (TREMBLrel_02, Created)		
DT	01-FEB-1997 (TREMBLrel_02, last sequence update)		
DT	01-MAR-2001 (TREMBLrel_16, last annotation update)		
DE	PURATIVE 98 KDA OUTER MEMBRANE PROTEIN.		
OS	Chlamydia psittaci (Chlamydophila psittaci).		
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.		
OX	NCBI_TaxID=83554;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-OVINE ABORTION S26/3;		
RA	Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;		
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: U72499; AAB18188.1; -		
DR	InterPro: IPR003357; -		
DR	InterPro: IPR003368; -		
DR	Pfam: PF02385; OMP: 1.		
DR	Pfam: PF02415; DUF145: 1.		
SQ	SEQUENCE 926 AA; 98439 MW; 3E755E52F594750F CRC64;		

```
Query Match      37.7%   Score 1802; DB 2; length 926;
Best Local Similarity    40.8%   Pred. No. 2,6e+98;
Matches          385; Conservative 172; Mismatches 354; Indels 32; Gaps 15
```

Oy      1 MKTSPWLVSVLA--PSCHQSLANFELPPDDSFNGINDSGFTPTKST-----YS 54  
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db      1 MPSPYKLIISETLTPISFHSQHLAEVALTOESILDAN---GAFSPDSTAGGTIVN 57

Qy	55	LTGDVEFEPECKGPIPLSSCCKOTDMLTEFLGNHSLTFPEGIDAGTAGAAGAASTFANKL	114
Db	58	VESDLSIVDVCQTAAALSSAAYVQADNLTEGNNHSLSTINNAAGAPAGIIVNTADKIL	117
Qy	115	TFSGFSLSPGSPSTVYVTTGQGLTSSAGVNLNIRKLVVAGNFSTADGAIRKASFL	174
Db	118	TLTDEPSKLFKEPCESLSIVNTGKAMKSGGALINNNASILEFDQYSAENGAISCKRAFSL	177
Qy	175	TGTSDDALFSSNNSSSTKGALATATAGARIANNQGYVAFSLNASTSGALIDDEGTSILSN	234
Db	178	TGSSKEISFTTNSTAKKGALAAAGIALHLSNOQSTIFPSGTAVNASGGAAYSEASMTIAG	237
Qy	235	NKFLTFEGNNAKTT---GGAICWTKASQSELLISNNKTLIFASNAVEYSGGAIAHAKL	290
Db	238	NNHVAFSNNAAVSGSDGGGAIHCKSTGSAPTLTIKDNKYLIFEEMTSAAKGAIYTDKL	297
Qy	291	ALSSCGFEPTLRNNVSSATPKGGAISIDASELSLSEGTNITFVRNTLTGTSQTDTPKR	350
Db	298	ILTSQGPAPFNNVYVHTAHTPKGAGIAGIANECESLTAHEGDTI-DNNLMAQDNATIKR	356
Qy	351	NAININGSGKTELRRAKNHITFYDDPTDTSQTSVDLKTINNSAGALNPQGTILFSGE	410
Db	357	NAINIEGKKNVNLRAASGKTISETDITYEBGNADLLITLKBAGD--KTYNGRIIFSGE	414
Qy	411	TLTADKLVAADNLKSSFTQPYSLSGKLLLOKQYLETBSTFSESQAGSLAMDGSTLLST	470
Db	415	KLTREQAAVADNLKTFPTQPTTLAAGELYVRSGEVEAKTYVQAGSLIMDAGTKLSAK	474
Qy	471	AGSTITTNLGINVDSLKQPVSLTAGASKNVYVSGKMLDIEGIVYSHMFHSDQLE	530
Db	475	TEDATLNLINAPMTLLOGKRAVDAVDAAGANNTLISAIGVIDPTGFEYENHLNPTLAL	534
Qy	531	SLKLTIVADADVDTVNDLSSILPVAEPDENSEYGFQOGMANNMTDITATNKKEATA--TWT	588
Db	535	GGIOLSGKGV--TTTNVPSHYVGAE---THYGQGMWSYMWKDNNSDPKQTQALFTWN	590
Qy	569	KTEGYVSPERKSALVCNTLKGVPFDITRSLQOLYEIGATG--MEHKQGFWWSMNFPLAKTG	647
Db	591	KTGYPNPERAPYPLNLSLWOSFIDLSIDYVERSDSTILETRRGILWMSGINFEFKOR	650
Qy	648	D-ENNKGGFRRHSGGVYGGASHPKDQLFFAFCHELPARKDCEFIANNSTRYQGTLEFK	706
Db	651	MAENRK--FRHISSTGYLGATNTSRBESLSVAPQOLFKAOKDVLVSKMANVYAGSVYQ	709
Qy	707	HSHTLOPONYLRILGRAKFSESAIEKFPREIPLALDVQVSFSSHDNRMEHTYSLPSEGS	766
Db	710	HYSKFDDTLRLFLNG----PNVCCSGSEKKEIDPIFLDAQITCHANNMTTSYTPYEVKGS	765
Qy	767	WSNRCIAGGTGLDLPF-VLSNPHLEKFTPIPOKKEVNVVYSSONSFPSSSDGGSIGRL	825
Db	766	WGNPTLGLTLTSYPIYFSS--SIFDSIAPAKQIYVAAHQDFEKPTEGKVFESSDL	823
Qy	826	LNLSIPVAKAFVQGDIDSDSYTDLSGFVSDVYRRNPNOSTATLWSPDSSKIRGCNLSRO	885
Db	824	LNVSAPIGIKFEKLSYGERSAYDITLAIYIDPVYRHNPSCKMTGLAINDVSWLTATMLARO	883
Qy	886	AFILRGSNNYVNSNCELFGHYAMELRLGSSRRNVYDGYTLRF	928
Db	884	AFIVRAGNHIALTSGVEMFQFGEGLSSSSRRNVNDVIGAVAR	926

RESULT	4
Q9Z398	
ID	Q9Z398
AC	Q9Z398;
DT	01-MAY-1999 (TrEMBLrel. 10, Created)
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT	01-MAR-2001 (TrEMBLrel. 15, Last annotation update)
DE	OUTER MEMBRANE PROTEIN PRECURSOR (POLYMORPHIC MEMBRANE PROTEIN G
DE	FAMILY).
GN	OMP10 OR PMP_9 OR CP0306.
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).
OC	Bacteriia: Chlamydiales: Chlamydiaceae: Chlamydia.

RT from Japan and CMI029 from USA.  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL: AJ001311; CA04672.1; -  
 DR EMBL: AJ133034; CAB37072.1; -  
 DR EMBL: AE001628; AAD18593.1; -  
 DR EMBL: AF002546; BAA98658.1; -  
 DR InterPro: IPR003357; -  
 DR InterPro: IPR003368; -  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.  
 KW signal.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT CHAIN 18 928 OUTER MEMBRANE PROTEIN 4.  
 SQ SEQUENCE 928 AA; 98903 MW; 788BCDD62C911402 CRC64;

Query Match 100.0%; Score 4782; DB 2; Length 928;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-274;  
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTSSIPWLVSVYLAFCCHQSLANEELSPDDSFNGNIDSGFTPTKTSATYSLTGDFV 60  
 DB 1 MKTSSIPWLVSVYLAFCCHQSLANEELSPDDSFNGNIDSGFTPTKTSATYSLTGDFV 60  
 QY 61 FYEPGKGTPLSDSCFKQTTDNLTFGLNGHSLTFGFDAGTHAGAASTANKMLTFESGFS 120  
 DB 61 FYEPGKGTPLSDSCFKQTTDNLTFGLNGHSLTFGFDAGTHAGAASTANKMLTFESGFS 120  
 QY 121 LLSFDDSPSTVTYTGQGTSSAGGVNLENIRKLVAGNFTAGGAIAKGSFLLTGSGD 180  
 DB 121 LLSFDDSPSTVTYTGQGTSSAGGVNLENIRKLVAGNFTAGGAIAKGSFLLTGSGD 180  
 QY 181 ALFSNNSSSTKGAIAITTAGARIANNNGYRFLSNIASTSGAIDDEGTIILSNKKLYF 240  
 DB 181 ALFSNNSSSTKGAIAITTAGARIANNNGYRFLSNIASTSGAIDDEGTIILSNKKLYF 240  
 QY 241 EGNAAKTGGAICNTRKASGSPPELLISNNKTLIFASNAVETSGAIAHAKKIALSSGGFTEF 300  
 DB 241 EGNAAKTGGAICNTRKASGSPPELLISNNKTLIFASNAVETSGAIAHAKKIALSSGGFTEF 300  
 QY 301 LRNNVSATPKGAISTDAGSELSSAETGNITFVRNTLTGTTSTGTPKRNAINIGSNGK 360  
 DB 301 LRNNVSATPKGAISTDAGSELSSAETGNITFVRNTLTGTTSTGTPKRNAINIGSNGK 360  
 QY 361 FTTELRAKNNHTIFFYDPTSEGTSSDVLKINNAGALNFPYOGTILFSGEFTLTADELKVA 420  
 DB 361 FTTELRAKNNHTIFFYDPTSEGTSSDVLKINNAGALNFPYOGTILFSGEFTLTADELKVA 420  
 QY 421 DNLKSTFQPVSLSGKLLQKGVLTLESTFSQDAGSLGMDSGTLLSTTAGSITTTNLG 480  
 DB 421 DNLKSTFQPVSLSGKLLQKGVLTLESTFSQDAGSLGMDSGTLLSTTAGSITTTNLG 480  
 QY 481 INVDISGLKQPVSLTAKGANKYVSGKLLIDIEGITYSHMFSDQLFSLKITYDAD 540  
 DB 481 INVDISGLKQPVSLTAKGANKYVSGKLLIDIEGITYSHMFSDQLFSLKITYDAD 540  
 QY 541 VDNVNDISSILIPYPAEDPNSEYFGOGOMNVMNTDTATNKEATATWTKGFYSPERKS 600  
 DB 541 VDNVNDISSILIPYPAEDPNSEYFGOGOMNVMNTDTATNKEATATWTKGFYSPERKS 600  
 QY 601 ALVYCNITMGVFTDIRSLQQLVEIGATGMEHKQGFVWSMTNPLHKTGDENRKGFRRHSGG 660  
 DB 601 ALVYCNITMGVFTDIRSLQQLVEIGATGMEHKQGFVWSMTNPLHKTGDENRKGFRRHSGG 660  
 QY 661 YVIGGSAHTEKDLTFPAFCFLFARPKDCFLIANNSTRVGTLEFFHSHSLQONLRIG 720  
 DB 661 YVIGGSAHTEKDLTFPAFCFLFARPKDCFLIANNSTRVGTLEFFHSHSLQONLRIG 720  
 QY 721 RAKFSESAIEKFPREIPLALDVVSFSDNRMEHTYSLPESEGSMSNCEIAGGIGLID 780  
 DB 721 RAKFSESAIEKFPREIPLALDVVSFSDNRMEHTYSLPESEGSMSNCEIAGGIGLID 780  
 QY 781 PFVLSNHPLEKFTIFOMKVEMYVVSQNSFFESSSDGRGFSIGRLNLSTIPVAKFVQGD 840

DB 781 PFVLSNHPLEKFTIFOMKVEMYVVSQNSFFESSSDGRGFSIGRLNLSTIPVAKFVQGD 840  
 QY 841 IGSYTYDLSGFFVSDYRRNNPOSTATLVMSPPSMKIRGNLSRQARLLGSSNNYYNSN 900  
 DB 841 IGSYTYDLSGFFVSDYRRNNPOSTATLVMSPPSMKIRGNLSRQARLLGSSNNYYNSN 900  
 QY 901 CELFGHYAMELRGSSRRNYNDVGTKLRF 928  
 DB 901 CELFGHYAMELRGSSRRNYNDVGTKLRF 928

RESULT 2  
 Q9K299 PRELIMINARY; PRT; 949 AA.  
 AC Q9K299;  
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)  
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)  
 DE 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)  
 DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.  
 GN CP0302.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83358;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39.  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umayan L.A., Uterback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 pneumoniae AR39";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AE002192; AAF38159.1; -  
 DR TIGR: CP0302; -  
 DR InterPro: IPR003357; -  
 DR InterPro: IPR003368; -  
 DR Pfam: PF02385; OMP; 1.  
 DR Pfam: PF02415; DUF145; 1.  
 SQ SEQUENCE 949 AA; 101357 MW; A00B09E16C699BE3 CRC64;

Query Match 100.0%; Score 4782; DB 2; Length 949;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-274;  
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTSSIPWLVSVYLAFCCHQSLANEELSPDDSFNGNIDSGFTPTKTSATYSLTGDFV 60  
 DB 22 MKTSSIPWLVSVYLAFCCHQSLANEELSPDDSFNGNIDSGFTPTKTSATYSLTGDFV 81  
 QY 61 FYEPGKGTPLSDSCFKQTTDNLTFGLNGHSLTFGFDAGTHAGAASTANKMLTFESGFS 120  
 DB 82 FYEPGKGTPLSDSCFKQTTDNLTFGLNGHSLTFGFDAGTHAGAASTANKMLTFESGFS 141  
 QY 121 LLSFDDSPSTVTYTGQGTSSAGGVNLENIRKLVAGNFTAGGAIAKGSFLLTGSGD 180  
 DB 142 LLSFDDSPSTVTYTGQGTSSAGGVNLENIRKLVAGNFTAGGAIAKGSFLLTGSGD 201  
 QY 181 ALFSNNSSSTKGAIAITTAGARIANNNGYRFLSNIASTSGAIDDEGTIILSNKKLYF 240  
 DB 202 ALFSNNSSSTKGAIAITTAGARIANNNGYRFLSNIASTSGAIDDEGTIILSNKKLYF 261  
 QY 241 EGNAAKTGGAICNTRKASGSPPELLISNNKTLIFASNAVETSGAIAHAKKIALSSGGFTEF 300  
 DB 262 EGNAAKTGGAICNTRKASGSPPELLISNNKTLIFASNAVETSGAIAHAKKIALSSGGFTEF 321  
 QY 301 LRNNVSATPKGAISTDAGSELSSAETGNITFVRNTLTGTTSTGTPKRNAINIGSNGK 360  
 DB 322 LRNNVSATPKGAISTDAGSELSSAETGNITFVRNTLTGTTSTGTPKRNAINIGSNGK 381

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2001, 03:27:00 ; Search time 119.05 Seconds  
(without alignments)  
1031.324 Million cell updates/sec

Title: US-09-446-677b-2  
Perfect score: 4782  
Sequence: 1 MKTSTPWWLVSSVLAFFSCHL.....MELRGSSRRNYNDVGTKLRF 928

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_16:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_unclassified:\*  
13: sp\_vertebrate:\*  
14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4782	100.0	928	2	086164
2	4782	100.0	949	2	09K299
3	1802	37.7	926	2	P71135
4	1794	37.5	928	2	092398
5	1785	37.3	928	2	09RB65
6	1763	36.9	930	2	09RB66
7	1761	36.8	930	2	092393
8	1739	36.4	914	2	086163
9	1734	36.3	936	2	092898
10	1733	36.2	936	2	09J842
11	1591	33.3	772	2	09RB71
12	1562.5	32.7	839	2	P77792
13	1535	32.1	846	2	P71133
14	1532.5	32.0	847	2	P71132
15	1435.5	30.0	841	2	092381
16	1353	28.3	922	2	092965
17	1352	28.3	922	2	09K1Y9
18	1344	28.1	922	2	0924H9
19	1284.5	26.9	973	2	092896

20	1284.5	26.9	995	2	09K2A1	09K2A1 chlamydia p
21	1203	25.2	1276	2	09JRM2	09JRM2 chlamydia p
22	1164	24.3	712	2	09RB73	09RB73 chlamydia p
23	1144.5	23.9	1407	2	092899	092899 chlamydia p
24	1102.5	23.1	1013	2	084879	084879 chlamydia m
25	1040	21.7	987	2	09PL45	09PL45 chlamydia t
26	843.5	17.6	494	2	09RB68	09RB68 chlamydia p
27	837	17.5	878	2	084882	084882 chlamydia t
28	831.5	17.4	427	2	09RB70	09RB70 chlamydia p
29	830.5	17.3	649	2	P71134	P71134 chlamydia p
30	829.5	17.3	867	2	09PL41	09PL41 chlamydia m
31	790	16.5	445	2	09RB67	09RB67 chlamydia p
32	736.5	15.4	359	2	09JSK6	09JSK6 chlamydia p
33	679.5	14.2	947	2	092813	092813 chlamydia p
34	678.5	14.2	947	2	09JSR2	09JSR2 chlamydia p
35	610.5	12.8	1723	2	092812	092812 chlamydia p
36	610.5	12.8	1723	2	09RB59	09RB59 chlamydia p
37	610.5	12.8	1732	2	09K2C1	09K2C1 chlamydia p
38	608.5	12.7	978	2	09RB63	09RB63 chlamydia p
39	604.5	12.6	986	2	09PJV3	09PJV3 chlamydia m
40	604	12.6	1016	2	084880	084880 chlamydia t
41	603.5	12.6	978	2	092895	092895 chlamydia p
42	600.5	12.6	934	2	09JSE7	09JSE7 chlamydia p
43	599.5	12.5	934	2	092882	092882 chlamydia p
44	599.5	12.5	952	2	09K2A5	09K2A5 chlamydia p
45	590	12.3	946	2	092880	092880 chlamydia p

## ALIGNMENTS

## RESULT 1

ID 086164 PRELIMINARY: PRT: 928 AA.

AC 086164;  
DT 01-NOV-1998 (TRENBLER, 08, Created)  
DT 01-NOV-1998 (TRENBLER, 08, Last sequence update)  
DT 01-MAR-2001 (TRENBLER, 16, Last annotation update)  
DE OUTER MEMBRANE PROTEIN 4 PRECURSOR.  
GN OMP4 OR Pmp\_11.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC/CWL-029/VR-1310.  
RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;  
RT "Identification of two novel genes encoding outer membrane complex  
associated surface layer proteins in Chlamydia pneumoniae.";  
RL submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VR1310;  
RA Hierno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,  
RT Christiansen G., Birkelund S.;  
RL "Chlamydia proteins containing the GGA-repeat belong to a subfamily  
of autotransporting pathogenicity factors.";  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWL029;  
RA MEDLINE=99206606; PubMed=10192388;  
RX Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RL "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RA MEDLINE=20330349; PubMed=10871362;  
RX Shira M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RL "Comparison of whole genome sequences of Chlamydia pneumoniae J138

```

Db      1 mkpilhllisctlvprllslatygadaslspcdsfdgags--tftpxstadangtnyv 59
QY      59 LMGDITLQNLGDSALATKGCFSPTTESLSPAGKYSLSPLNIK-SAEAGALSVTTDKNL 117
Db      60 lsgnyvndagkgtaltgccttctgtlfttgykysfntvdagsnagaasttdkal 119
QY      118 SLTGFSLLFLAAPSIVTTPSGKAVKCGDLTFPNNGTILFKODYCEE---NGAIST 174
Db      120 tftgfentlsiaapgt--tvaagkctlsesagalnldngtlltsqnvseannngaait 177
QY      175 KNLSTKNSGTSIFEGKSSATGKGAICATGTVDITNTAPTLFSNNIAEAGALINS 234
Db      178 ktlsisgnstslftsnsak---klgalyssaaslsqtlgtvfmnkygetcgagalf 234
QY      235 TGNCTITGNTSLVFSNSVTATANGAL---SGDA-DVTISGNOSVTFSSQAOVANG 289
Db      235 eaassltqnslltsfsgntdaagkgaicyektgetptlltsqnsltfaenssvtqg 294
QY      290 AIYAKKLTLASGGGCGISFSNNIVGCTTAGNGAISTILAECSLSAEAGDITFNGNAIV 349
Db      295 aicahgldisa--apgtlismnrcgntaagkgajaladsgslsangdltflgnlt 352
QY      350 ATPPTTKRNSIDIGSTAKITNLRAISGHSIFPYDPTANTAADSTDTLNLKADAGNST 409
Db      353 staptstrnailygsakltlnraaggslyfydpilasntg--asdvltlmpdsnpl 411
QY      410 DYSGSIVFSEKERISDEEAVADNLSTLKQPVTLTAGNLVLKRGVLLDRTGFTQTSV 469
Db      412 dyagvtfvgekelssadeaadaftslkxpialasgalknveldvngtqgtqestl 471
QY      470 IMAGTTLKASFEETVLTGSLIPVDSLGEKKVYIAASAASKNVALSGPILLDNOGNAY 529
Db      472 lmpgkkladetaisltklvslaslegnkysietaganiktltspvfdsgnfy 531
QY      530 ENHDLCKTODESFVQLSALGTATTT-----DYPAPVYATP-THYGYOCTWMTWDDT 562
Db      532 eshtlntg---ftgplvfvftaataasdiydalltsprqtphephygyqhwatwad-- 585
QY      583 ASPPKTKTATLMTNTGTYLPNPERGRLVNPNSLMSGFSIOALOGVENSALTLCSDRGT 642
Db      586 ---lstksgcmvctvgypnperrtasvvpdsiwasftdrtltqtlmcsqanslyqrgl 643
QY      643 WAAGVANFLDKDKGKKRYRHRKSGGYAIGAAGTCESENLISFAFQOLFGSDKDFLVAKN 702
Db      644 waagtanftfkdksgntqatfrhksygyivgsssedfsemlfsafqllfgkdkdlfven 703
QY      703 HDTYAGAFYIQTITRCSEF---IGCLLDKLPGSWSHKRPLVEGQLANSVSHVNDIKTKY 758
Db      704 tshnylaslylqtratlqglpmpsfgsitdml---kdipllinaqslsytkndmdtry 759
QY      759 TAYPEKSGSNGNAFMMNLGASSHSY-PEYLHCFEDTYAPYIKNLVYIRNDSSEKTEG 817
Db      760 tsypesgsvtnmsgalelgyslaalykpeapfifgyfplkfgavysrktkesga 819
QY      818 RSPDSDNLNLSLPIGVKPEKESDCNDFSYDLTSLVYDPIRNDPKCTVALVIGASWET 877
Db      820 ratddgdvlncspvgrlreksiedekmfelslaylgtvyrknprsrtslmsgswts 879
QY      878 YANNLARQALQVRAGSHYAFSPFEVLGQFVEVRGSSRIYVNDLGKRFQF 928
Db      880 lcnlarqelaasaghltslphvelsgeaayelrgsahlnvdcgllrysf 930

```

## RESULT 11

AA35052 standard; Protein; 930 AA.

AC AA35052;

XX 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae surface exposed polypeptide.

```

XX      KW Respiratory disease: pneumonia; bronchitis; heart disease; sarcoidosis;
KW      sinustitis; purulent otitis media; erythema nodosum; pharyngitis;
KW      vaccine; neutralising epitope.
XX      OS Chlamydia pneumoniae.
XX      PN MO9927105-A2.
XX      PD 03-JUN-1999.
XX      PF 20-NOV-1998; 98WO-1B01890.
XX      PR 04-NOV-1998; 98US-0107078.
XX      PR 21-NOV-1997; 97FR-0014673.
XX      PA (BEST ) GENSET.
XX      PI Griffals R;
XX      PI WPI; 1999-357842/30.
XX      PT Genome sequence of Chlamydia pneumoniae
XX      PS Page 940-942; Disclosure; 1912pp; English.
XX      CC AA34584-Y35879 represent the proteins encoded by all the open reading
CC      frames in the complete genome (see AA391990) of Chlamydia pneumoniae.
CC      CC pneumoniae causes respiratory disease such as pneumonia and
CC      bronchitis and is thought to be a contributing factor in heart
CC      disease, sarcoidosis, sinustitis, purulent otitis media, erythema
CC      nodosum or pharyngitis. The polypeptides encoded by the open reading
CC      frames of the C. pneumoniae genome (see AA34584-Y35879) can be used in
CC      immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC      nucleotide sequences can also be used as immunogenic compositions,
CC      especially where the vector directs the expression of a neutralising
CC      epitope of C. pneumoniae.
XX      SO Sequence 930 AA:

```

Query Match 42.3%; Score 2028; DB 20; Length 930;

Best local similarity 44.5%; Pred. No. 4,6e-125;

Matches 423; Conservative 158; Mismatches 326; Indels 44; Gaps 17;

```

QY      1 MKQGFMLVLSSTLACTSCSTVFATAFENIGSPDSFSTNTGYTPKNT--TTGIDY 58
Db      1 mkpilhllisctlvprllslatygadaslspcdsfdgags--tftpxstadangtnyv 59
QY      59 LMGDITLQNLGDSALATKGCFSPTTESLSPAGKYSLSPLNIK-SAEAGALSVTTDKNL 117
Db      60 lsgnyvndagkgtaltgccttctgtlfttgykysfntvdagsnagaasttdkal 119
QY      118 SLTGFSLLFLAAPSIVTTPSGKAVKCGDLTFPNNGTILFKODYCEE---NGAIST 174
Db      120 tftgfentlsiaapgt--tvaagkctlsesagalnldngtlltsqnvseannngaait 177
QY      175 KNLSTKNSGTSIFEGKSSATGKGAICATGTVDITNTAPTLFSNNIAEAGALINS 234
Db      178 ktlsisgnstslftsnsak---klgalyssaaslsqtlgtvfmnkygetcgagalf 234
QY      235 TGNCTITGNTSLVFSNSVTATANGAL---SGDA-DVTISGNOSVTFSSQAOVANG 289
Db      235 eaassltqnslltsfsgntdaagkgaicyektgetptlltsqnsltfaenssvtqg 294
QY      290 AIYAKKLTLASGGGCGISFSNNIVGCTTAGNGAISTILAECSLSAEAGDITFNGNAIV 349
Db      295 aicahgldisa--apgtlismnrcgntaagkgajaladsgslsangdltflgnlt 352
QY      350 ATPPTTKRNSIDIGSTAKITNLRAISGHSIFPYDPTANTAADSTDTLNLKADAGNST 409
Db      353 staptstrnailygsakltlnraaggslyfydpilasntg--asdvltlmpdsnpl 411

```





PT Chlamydia pneumoniae - comprises detecting antibodies specific for  
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding  
 these proteins  
 XX  
 PS Claim 7: Page 53-55; 115pp; English.  
 CC This polypeptide comprises the novel 90.0 kDa surface exposed  
 CC protein Omp6 of the human respiratory pathogen Chlamydia  
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see  
 CC AA006820) isolated from a C. pneumoniae expression library. The  
 CC invention provides 12 novel surface exposed proteins, Omp4-Omp13  
 CC (see AA006817-28), and nucleic acid sequences encoding them (see  
 CC AA006816-27). A new species specific test is claimed that is used  
 CC to identify mammals (including humans) infected with Chlamydia  
 CC pneumoniae. The test comprises detecting antibodies specific for  
 CC Omp4-Omp13 or detecting nucleic acid fragments encoding these outer  
 CC membrane proteins, especially by PCR. The proteins are also used  
 CC in the diagnosis of C. pneumoniae infection in mammals. The  
 CC nucleic acids and proteins can also be used in the immunization of  
 CC mammals, the nucleic acids being particularly useful as DNA  
 CC vaccines for effecting in vivo expression of antigens. The  
 CC vaccines may also prevent atherosclerosis and bronchial asthma,  
 CC which are possibly associated with C. pneumoniae.

XX Sequence 928 AA:

Query Match 43.1%; Score 2069; DB 20; Length 928;

Best Local Similarity 46.9%; Pred. No. 9.2e-128;

Matches 446; Conservative 141; Mismatches 319; Indels 44; Gaps 20;

QY 1 MKQSFMLVLSSTLACFTSCSTVFATAENIGPSDFSGSTNTGTVPKMT--TTGIDYT 58  
 DB 1 mkxsfkfvf-stfalf--plsmiatekvldssaaafdgkn-gnfsvrsqgedagtl 55  
 QY 59 LFGDITLQNL-GDSALTKGCFSDTETESLPACKGYSFLNLS-SAEQAL-SVTTDK 115  
 DB 56 fkgvntlemipgftaltkcfntkgtlftgngnallfgtvdagvagaavnsavdk 115  
 QY 116 NLSTGFSSTLFLAAPSVTTPSGKAVKCG-GDLTFQNNGTLTFQDCENGAIST 174  
 DB 116 stftgfsstlsfapsvtltpsgkavkcg-gdltfqnnngtltfqodcengais 173  
 QY 175 KNLKNSGTSGISPEGNKSSATGKGCALCATGTVDITNTTAFLPSNNIAEAGAINS 234  
 DB 174 klslsgtmsalfsents---kkgalqtsdalitltgqgevsfdntsssgaaitf 230  
 QY 235 TGCCTTGNTSLVSENSVT---ATAG--NGAL-----SGDADVTISGNQSVTESG 283  
 DB 231 eaavtlnnakvafidnkvgaassstltdmsggaicayktsdtkvltltnqmlf 280  
 QY 284 AVANGAIYAKKLTLLASGGGGISFSNNIVQGTAGNGAISLILAGECSLSAEGDIT 343  
 DB 291 straggaikykkleias--ggaltfsvngtclapkgalaiedsgelsdsdgdlvf 348  
 QY 344 NGNAIVATPQTKRNSIDIGSTAKITNLRAISGHSIFFPDPTTANTADSTTLNKA 403  
 DB 349 lgnltvstlp-gfnrssiidgtlsakmlrtaagraiyfdrplltgtsltvdlk 407  
 QY 404 DAGNSIDYSSSIYFSGEKLSDEAKYADNLTSLKQPVTLTLAGNLVLRKVTLDTKG 463  
 DB 408 pdsasalytgnliltfgeklssetaasknltsklqrvlsggltlskhlqvtqatfg 467  
 QY 464 TACSSIVMDGTTLKASTEEVTLTGSLIPVDSIGEGKRVVIASASAKNVALSPILL 533  
 DB 468 qadsrlemdvgttlepa-dstlnmlvlnissldgakkakietkasknltlsgtllid 526  
 QY 524 NQGNAYENHDLGATQDFSVQLSALGATATTDPVAVPTVATPHYGYGTG--MTW 582  
 DB 527 plgtfyenhlrmpgsydllelkasgvtstavlppdlmgekfhygygtwpgl 582  
 QY 583 ASPPKYTAALANTNGCYLNPERRGGLVYVNSLMGFSFDLQALQGVTERKALALSC 642  
 DB 583 ASPPKYTAALANTNGCYLNPERRGGLVYVNSLMGFSFDLQALQGVTERKALALSC 642

DB 583 -gtgaatlafnwkltkyipnerigs\lvpslwnafidisslhy\metaneg\lqgrat 641  
 QY 643 WAAGVANFLDKDKKGERKRRHKSAGGAIGCAOTCSENLSFAFCOLFGSDDFLVAKN 702  
 DB 642 wcaglsnifhkdsktkrtgfrhlsygnlnhtcsdkllsaefqlfrdrdyfrakn 701  
 QY 703 HTDTYACAFYIQH---ITPESGFICGLDLKPLPGSWSHKPLVLEGOLAVHSVNDLTK 758  
 DB 702 qgtvgygltlyyghneylslpcklrpslslyvp---telvlsnglnsyhtndldtky 758  
 QY 759 TAYPEVKGSGNNAPNMLGASSHSITPEYLACDPTAPYTKLNLTYRKDSFSKTEGR 818  
 DB 759 tlyptvgsygnsdafalefggrapicdesaleqymplnkltfyahgegfkegrear 818  
 QY 819 SPFDSNLFNLSTPLGVKEFSPONDESYDLTSLYVDLIRNPCKTALVIGASMETY 878  
 DB 819 efgssrtvmlalprgfrldescqcdatyltlgytvdlyvrsnmpdctllrlsgds 878  
 QY 879 ANNILARQALOVRAGSYAFSPMEVLGQFVEYVGSRIYVNDLGKRFQF 928  
 DB 879 gtnlarqalvrlagnhctfnsteafsfelrgrsrynvudlgakyqf 928

RESULT 9

AA94327  
 ID AA94327 standard; Protein; 928 AA.

XX AA94327;

XX 11-AUG-2000 (first entry)

XX Chlamydia pneumoniae 98kD putative outer membrane protein.

XX Chlamydia; antigen; vaccine; infection; outer membrane protein.

XX Chlamydia pneumoniae.

XX WO200026237-A2.

XX 11-MAY-2000.

XX 29-OCT-1999; 99MO-GB03579.

XX 29-OCT-1998; 98US-0106070.

XX 01-MAR-1999; 99US-0122066.

XX 27-OCT-1999; 99US-0428122.

XX (CONN-) CONNUGHT LAB LTD.

XX Mordin AD, Oomen RP, Dunn PL;

XX WPI: 2000-365569/31.

XX N-PSDB: AAA27021.

XX Novel Chlamydia 98 kDa putative outer membrane protein antigen, used  
 PT for vaccination and protection against Chlamydia infection -  
 PS Claim 6: Fig 1: 93pp; English.

XX The present sequence is the 98kDa putative outer membrane protein from  
 CC Chlamydia pneumoniae. The genomic sequence was amplified using two  
 CC PCR primers. The 5' primer contains a NotI restriction site, a ribosome  
 CC binding site, an initiation codon and a sequence close to the 5' end of  
 CC the 98kDa putative outer membrane protein coding sequence. The 3' primer  
 CC contains the sequence encoding the C-terminal sequence of the putative  
 CC outer membrane protein and a BstXI restriction site. The stop codon was  
 CC excluded and an additional nucleotide was inserted to obtain an in-frame  
 CC C-terminal fusion with the Histidine tag. The PCR product was cloned  
 CC into a eukaryotic expression vector (pCA-Myc-His) by performing both  
 CC the vector and the PCR product with NotI and BamHI and performing a  
 CC ligation reaction. This expression vector was injected intramuscularly  
 CC and intranasally into mice, which were subsequently inoculated with  
 CC Chlamydia pneumoniae. The chlamydial lung titers of the immunised mice

```

XX 30-DEC-1998.
PD 19-JUN-1998; 98WO-DK00266.
XX 23-JUN-1997; 97DK-0000744.
PR (BIRK/) BIRKELUND S.
XX (CHRI/) CHRISTIANSEN G.
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mygind P;
XX WPI: 1999-105610/09.
DR N-PSDB; AAX06821.
XX
PT Species-specific test for identifying mammals infected with
PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT these proteins
XX
PS Claim 7; Page 56-58; 115pp: English.
XX
CC This polypeptide comprises the novel 96.7 kDa surface exposed
CC protein Omp9 of the human respiratory pathogen Chlamydia
CC pneumoniae. Its amino acid sequence was deduced from DNA (see
CC AAX06821) isolated from a C. pneumoniae expression library. The
CC invention provides 12 novel surface exposed proteins, Omp4-Omp15
CC (see AAW8417-28), and nucleic acid sequences encoding them (see
CC AAX06816-27). A new species specific test is claimed that is used
CC to identify mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.
CC
XX Sequence 918 AA:
SQ

```

```

DB 347 tsttspasvtrnaidlsnackfnlratrgukvlfydpilssga---tdkrlskadag 403
QY 407 NSTDYSGSIVFSGEKLSEDEKAVADNLSTLKQPTTLRAGNLVLRGYTLDTKGTQTAG 466
DB 404 sgnlyegylvisgeklsseeelkphnlkfstfqqavelaagaalvlkdgytvaantltyeg 463
QY 467 SSVIMDAGTTLKASPEEYTLTGLSPVDSLGEKGVVVAASAASKVNLSPRIILLDNQ 526
DB 464 skvmdggctfcaaseayrlnglalnidsldgtakalkakaasakvalsgpmlvdaag 523
QY 527 NAYENHDLGKTQDFSEFVOLSALGATTTDVPAVPVATPTHYGVGTGWTWDDTASTP 586
DB 524 nyehhnlsgqvfplielasagwtcttdpplntlnhygygctglvwdd--ata 581
QY 587 KTKTATLMTNTGTYLPNERGCPVLPNSLMSFSDIOALOGVIERASATLCSDRFMAAG 646
DB 582 ktknatltwtktgykphnperqplvpnsllwgsfydvrisqlmdtrstslssnllwsg 641
QY 647 VANFLDKDKGKRRKRRKSGGYAIGAAOJCSENLISFAFCQLGSDKDFLAKNHTDT 706
DB 642 iadllheqkqngqrsyrnsaagyalgg9fftaasenfifacqltygkdhvaknhbv 701
QY 707 YAGAFYIQTIECSGFTICLDKLPGSMHRPLYLEGOLATSVHSNDKTYTAYPEVG 766
DB 702 yagamsyrhlgesk---tlaklsgnsdslpfvfnarfayghdmmntktytgyvpg 757
QY 767 SWGNNAFNMATG-----ASSHSTPEYLHCEDTYAPIKLNLTYIRDSESEKTEERS 819
DB 758 swgnadafleggaipvvasgrsw-----vdhtplfnlemyahbnkengtegrs 811
QY 820 FDDSNLFLSLPIGVKFEKFSQCDNFSDYDLRLSYVPLIRNDPKCTRTALVSGASWERTYA 879
DB 812 fgsedlnlavpvgkikaktsdks--tydlslaypovirlnopgcttllmwsdswcsg 869
QY 880 NNLARQALQVRAGSHYAPSPFEVLGQVFPEYRGSSRIYNDLGKRFQ 928
DB 870 tslsrgallvragnhafanfevrsqfvelrgsrayaldigrfjf 918

```

RESULT 8  
AAW88421  
ID AAW88421 standard; Protein: 928 AA.  
XX  
AC AAW88421;  
DT 26-APR-1999 (first entry)  
XX  
DE Chlamydia pneumoniae surface exposed protein Omp8.  
XX  
KW Omp8; outer membrane protein 8; surface exposed protein; antigen;  
KW Infection; diagnosis; vaccine; atherosclerosis; asthma.  
XX  
OS Chlamydia pneumoniae.  
XX  
FN WO9858953-A2.  
XX  
PD 30-DEC-1998.  
XX  
PF 19-JUN-1998; 98WO-DK00266.  
XX  
PR 23-JUN-1997; 97DK-0000744.  
XX  
PA (BIRK/) BIRKELUND S.  
PA (CHRI/) CHRISTIANSEN G.  
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;  
PI Mygind P;  
XX  
DR WPI: 1999-105610/09.  
DR N-PSDB; AAX06820.  
XX  
PT Species-specific test for identifying mammals infected with

```

RESULT 6
ID AAY69369 standard; Protein; 918 AA.
XX
AC AAY69369;
XX
DT 19-JUN-2000 (first entry)
XX
DE Amino acid sequence of the CPN100395 polypeptide.
XX
KM CPN100395; Chlamydia infection; immune response; vaccine.
XX
OS Chlamydia pneumoniae.
XX
PN WO20001183-A2.
XX
PD 02-MAR-2000.
XX
PF 18-AUG-1999; 99WO-IB01449.
XX
PR 20-AUG-1998; 98US-0097187.
XX
PR 20-AUG-1998; 98US-0097189.
XX
PR 20-AUG-1998; 98US-0097189.
XX
PR 20-AUG-1998; 98US-0097190.
XX
PR 20-AUG-1998; 98US-0097195.
XX
PR 20-AUG-1998; 98US-0097196.
XX
PR 20-AUG-1998; 98US-0097197.
XX
PR 27-AUG-1998; 98US-0097191.
XX
PR 17-AUG-1999; 99US-0376770.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Murdin AD, Oomen RP;
XX
DR MPI; 2000-224703/19.
XX
DR N-PSDB; AAZ61509.
XX
PT Novel antigens and corresponding DNA molecules that can be used to
PT prevent, treat and diagnose disease caused by Chlamydia infection in
PT mammals, especially humans.
XX
PS Claim 19; Fig 15-E; 201pp; English.
XX
XX
AY69362-69 represent Chlamydia pneumoniae polypeptides. The
CC polypeptides are present in the bacterial membrane structure, in the
CC external vicinity of the membrane structure, in the inclusion membrane
CC structure, in the external vicinity of the inclusion membrane structure,
CC and in the cytoplasm of the infected cell. The polypeptides may be
CC used to prevent, treat and detect the presence of Chlamydia infection
CC and/or the presence of Chlamydia in a sample. The polypeptides may
CC also be used to induce an immune response in a mammal. The vaccine
CC vector comprising the polynucleotides is used to induce an immune
CC response in a mammal. Antibodies directed against the polypeptides
CC may also be used therapeutically to treat and/or prevent a Chlamydia
CC infection.
XX
SQ Sequence 918 AA;

```

```

Query Match 46.7%; Score 2237; DB 21; Length 918;
Best Local Similarity 49.5%; Pred. No. 8.3e-139;
Matches 470; Conservative 142; Mismatches 285; Indels 52; Gaps 18;

```

```

QY 1 MKSQFSLVLSSTLACTSCSYVPAATAF-NIPSDSFQDSTGTGYTPKNTT---TGID 56
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 mstfslilssla-fplmsvadaadltsrdsyngdltstftkaatsdsgtt 59
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 57 YTLGDTITLONLGDAAITKGCFSPTTESLSPAGKGYSLFL-IRSSAGALSTYTDK 115
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 60 yldgdvstsqgskqtsltscstntagnltfngfshfnlssstvagvvsntaas 119
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
PN

```

```

QY 116 NLS-LTGFSSLTFLAAPSIVTTPSGKAVKGGDLTFDNNGTLLFQKODYCEENGAI 174
DB ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 120 gtlkfsgfslmlaaprt---tgkgaikltgdlvstgnldlenassengaint 174
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 175 KNLKLNSTGSIPEFGKSSATGKGAICATGVDITNTAPTLEFNNIAEAGAGAINS 234
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 175 ktlslgsrtfvalignss---qggaiyasgdsvisenagllsfngnattsqgaisa 231
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 235 TGNCTTGNTSLVSENSVATAGNGAL-----SGDADVITSGQSTFSGNQAVAN 287
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 232 egnlvisnqnqnlffdgkact--nggaicdnkaganpdpdlitlsgneslhlnttngs 288
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 288 GGAIVAKKLTLASGGGGSIFSNINIVGTAGNGAISIILAAECSSAAGDITFGNA 347
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 289 ggaivakklvls-grgyvfsmkkaanap-kggaiaidsgeaisadgnlffegnt 346
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 348 IVAT-TPQTTKNSIDIGSTAKITNLRASGHISFEYDPIITANTADSTDLINKADAG 406
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 347 tsctspasvtrnaidiasnakflnlatrgnkvlfydpiltsaga---tdklslnkadag 403
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 407 NSTDYSGSIYFSGEKLSDEKAVADNLTSLKQPVYTLAGNLVKKRGVTLDTKFTQTAG 466
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 404 sgntlegylvifsgeliseelkkrpdlksltfgaveiaagalykkdgvlvvaantlvqeg 463
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 467 SSVIMDAGTTLKASTEVEVTLGLSIPVDSLEGKRVIAASAASKNVALSGPILLDNOG 526
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 464 skvymdgtlftesaagvltnglnainidsldgtlkaliktaaskdvaylsgplmlvdaag 523
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 527 NAYENHDLGKTFDSFVQVLSALGTATTDDVPAPVATPAPHYGQVGTGNTWDDTASTP 586
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 524 nyehhnlsgqvfpllelaagvltmttdltdpdlntcthygygwmvlvwwdd--ata 581
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 587 KTKTATLAWTNTGYLPPPEOGPLVPNSLWGSFSDIQAIGVIERSLTLCSRGRFAAG 646
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 582 klxnaltlwtcltqkprpexqgplvpnslvgsivdvdsislmdrscslssctnlvavsg 641
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 647 VANFLDKKRGKERRKYRHKSGAYALGAAGTQSENLSFACOLFGSKDFLVAKNHTDT 706
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 642 iadtlhedqngnrshsagaalggfftaasenfnfctqlfygdkhlyvaknltlv 701
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 707 YAGAFYIQHTBESGFGCLDKLPGSMKPLYLEQOLAYSHVSDLKRYTAPYVKG 766
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 702 yagamsyrhlgesk---tlaklsgnsdsipfinaftaygthdmmtkylgyspvkg 757
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 767 SMGNNAFNMMLG-----ASSHSYPEYLHCPTPYARYIKLNTLYIRQDSFSEKTEG 819
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 758 swgnadafgleggaipvvaagrtsw-----vdthpflnlemyahqndfkengtegrs 811
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 820 FDSNLEFNLSLPIGVKFEKPSDCNDESYDULTSVDPDLIRNDPKCTTALVSGASWETVA 879
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 812 fgedlfnlavpyikfeksdk--tydislayvpdivrindpogctllmvsdswatcg 869
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 880 NNARQALOVRASSHVAFSPMEFYLGOFVEVRGSSNLYVWDLGKKQF 928
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 870 tsisrqallvraqnhafeasnfefsqflevelrgsrsvayldigrfgr 918
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
RESULT 7
AAW88422 standard; Protein; 918 AA.
ID AAW88422
XX
AC AAW88422;
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp9.
XX
KM Omp9; outer membrane protein 9; surface exposed protein; antigen;
KM infection; diagnosis; vaccine; atherosclerosis; asthma.
XX
OS Chlamydia pneumoniae.
XX
PN WO9858953-A2.

```

OY	224	IDEAAGAGAINSGNCITGNTSTVSESESVTATAGAGALSGADVTIGSNGVTSNGN	283
Db	181	IAEAGGAINSGNCITGNTSTVSESESVTATAGAGALSGADVTIGSNGVTSNGN	240
OY	284	AAVANGCAIYAKKILTLTASGGGGISFSNNITVCGTTAANGAISTLAEGCSLSAADIIF	343
Db	241	AAVANGGAIYAKKILTLTASGGGGISFSNNITVCGTTAANGAISTLAEGCSLSAADIIF	300
OY	344	NGNAIVATTPOTTKRNSIDIGSTAKITLNLRAISGHSIFFYDPIITANTAADSTDTLNLNKA	403
Db	301	NGNAIVATTPOTTKRNSIDIGSTAKITLNLRAISGHSIFFYDPIITANTAADSTDTLNLNKA	360
OY	404	DAGNSTDYSGSYFSESEKLSDEDEAKADNLTSLKQPVTLTACNLVLRKGYTLDTGTGFQ	463
Db	361	DAGNSTDYSGSYFSESEKLSDEDEAKADNLTSLKQPVTLTACNLVLRKGYTLDTGTGFQ	420
OY	464	TAGSSVIMAGTTLTKASTEETLTGISTIPVDSIGECKKVIAASAASKVNLASGPILLD	523
Db	421	TAGSSVIMAGTTLTKASTEETLTGISTIPVDSIGECKKVIAASAASKVNLASGPILLD	480
OY	524	NOGNMAYENHDLEKTDQDFSVQLSALGTAATTDVPAVPVATPTHTHYOGTGWGTTWDDTA	583
Db	481	NOGNMAYENHDLEKTDQDFSVQLSALGTAATTDVPAVPVATPTHTHYOGTGWGTTWDDTA	540
OY	584	SPPKRTATLANTNGYLPNPEROGGLVNSLWGSFSDIOALOGVIERALITCSRGFM	643
Db	541	SPPKRTATLANTNGYLPNPEROGGLVNSLWGSFSDIOALOGVIERALITCSRGFM	600
OY	644	AAVGNFLDKDKGGEKKRYRHKSGGYAIGAOTCEENLISFAFCOLFSGDKFLVAKNH	703
Db	601	AAVGNFLDKDKGGEKKRYRHKSGGYAIGAOTCEENLISFAFCOLFSGDKFLVAKNH	660
OY	704	TDTVAGAFYIOHITTECSGFTGCLDLKLPGSWSHKPLVEGQLAYSHVSNDLKRTYATPE	763
Db	661	TDTVAGAFYIOHITTECSGFTGCLDLKLPGSWSHKPLVEGQLAYSHVSNDLKRTYATPE	720
OY	764	VGSMGNNAFNMGLASSHSPEYELHCPTTAPYIKLNLTYIRQDSFSEKGTGRSFDOS	823
Db	721	VGSMGNNAFNMGLASSHSPEYELHCPTTAPYIKLNLTYIRQDSFSEKGTGRSFDOS	780
OY	824	NFENSLIPIGVFEKPSDCNDPSYDLTSLVYPLINDPKCTALVTSGASWETVANNIA	883
Db	781	NFENSLIPIGVFEKPSDCNDPSYDLTSLVYPLINDPKCTALVTSGASWETVANNIA	840
OY	884	ROALQVIRAGSHYAFSPMEFVLQOFVEVRGSSRIYVNDLGKGFQ	928
Db	841	ROALQVIRAGSHYAFSPMEFVLQOFVEVRGSSRIYVNDLGKGFQ	885
RESULT 5			
ID	AAV35056	AAV35056 standard; Protein; 643 AA.	
AC	AAV35056;		
DT	13-SEP-1999	(first entry)	
DE	Chlamydia pneumoniae transmembrane protein sequence.		
KW	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis		
KW	sinusitis; purulent otitis media; erythema nodosum; pharyngitis;		
KW	vaccine; neutralising epitope.		
OS	Chlamydia pneumoniae.		
PN	W09927105-A2.		
PD	03-JUN-1999.		
PF	20-NOV-1998;	98WO-IB01890.	
RR	04-NOV-1998;	98US-0107078.	
RR	21-NOV-1997;	97FR-0014673.	

Query Match	Best Local Similarity	Score	DB	Length
Matches 628; Conservative	100.0%	3275;	20;	643;
	0;	Mismatches 207;	Indels	Gaps
				0;

  

Query	301	GGGGGSGFNNIVCGTTANGAISTLAEGCSLAENGDTFNGNAIVATPPTTKRMS	360
Db	16	ggggggsifsnmivgqtltagngaisllaegcsiaaegdlfngnaivatpqtikrns	75
Qy	361	IDIGSTAKTNNIRATISGSHIFFYDDPTTANTADSDTTLNLRKADAGNSTDYSGSIVFSGE	420
Db	76	idtgstaktlnlratsgshiffydpitantadstclnlnrkadagnstcdygsivfsg	135
Qy	421	KLSEDEAKYADNLTSTLKQPVTLTAGNLVLRKGVLTDRKFTQRTAGSSVYMDAGTTLRAS	480
Db	136	klsedeakvradnlstlklqpvcltaglnlylkrqvcldckgftqtagssvymdagttlkas	195
Qy	481	TEEVYLTIGISTVYDLSIGBEKKVYIAASAASKVVALSGPILLDNOGNATYENNDLCKTQDF	540
Db	196	teevltigtislsvdyslgeqkkyvllaasaaskvaalsgplllldngnayenhdlgktqdf	255
Qy	541	SPVOISALGTATTTTQVPAVPAVTFPHTKHGYOCTWGMVNDQTPASTPKRTATLANTNGY	600
Db	256	stpvqisalgatcttdvpaavpvauprhnyggcwgmwvddtascpkktatclawlny	315
Qy	601	LPNPEROGPLVENSJLWGSFSDIOAIGVIERGALTLCSDRGFMAAGVANFLDKRKGKRR	660
Db	316	lpnperogplvpnsjlgwgsfstdlgaigvieraaltlcsdrgfmaagvanfldkdkkgekr	375
Qy	661	KRRHNSGGAITGAAGTQCEENLISRAFCQLRESDXDFLAKNHTPTYACAFYIQHTTES	720
Db	376	kyrhnsygyaiggaatqceenlisaifcqlrfsdxdfflvaknhhttyagafiyqhntees	435
Qy	721	GTIGCLLDLPSPWSHKKPLVLEGOALAYSHVNDLTKTKTAYEPEVSGMGNNAFFNMMLGAS	780
Db	436	gtigclldlpspwswhkplvleqgalayshvndlctkyatayevksgwnafnmmlgas	495
Qy	781	SHSYPEYLCEFTYAPYIKLNTLYTRODSFSEKGEGRSFDSDNLFTNLPLIGVFEKES	840
Db	496	shsypeylhceftypyklnltytrodsfeskgegrsfdsdnlfnlslpivfeks	555
Qy	841	DCNDFSYDFTLSVVDLIRNDPKCTTALVISGASMEYANNNIARQALOVRAGSHAFSPM	900
Db	556	dcndfsydftlslayvdllrindpkcttalvisgaswetyanniarqalgyragshafsp	615
Qy	901	FEVILGQFVFEVAGSSRIYVNLGGRFOF	928
Db	616	fevilgqfvfevgsariyvndlgkfgf	643

Matches	913; Conservative	0; Mismatches	1; Indels	0; Gaps	0;
OY	1	MKSOFSMLVLSSTACTSCSTVFAATAEINIGPSDSFDGSGNTCTYPRKNTTGIDTIT	60		
Db	1	mksfswlvislactscstvfaataenigpdsfdgntctyprkntctgdyllc	60		
OY	61	GDITLQNLGDSAAALTKGCFSDTTESLSFAGKGYSLSPFLNKSSAEGAALSVTTDKNLSLT	120		
Db	61	gdtlqnlgsdaalatkcfdsdtteslsfagkyyslspflnkssageaalvttcknlslt	120		
OY	121	GFSSLTFLAAPSSTVITTPSGKAVKCGGDLTFDNNGTILRKDYCENGCAISFLNKSLK	180		
Db	121	gfssltflaapsvlttppsgkavkcggdltfdnngtllfkdyceengaisflknslk	180		
OY	181	NSGSISEEGKSSATGKGAICATGTVDITNTATPLTFSNNIAEAGAGINSTGCTI	240		
Db	181	nsgrisefegkssatgkkgacatgvtldtntaptltsnnlaeaaagainstgcti	240		
OY	241	TGNTSLVSENSVATATANGGALSGADVATISGNQSVTFSGNQAVANGCAIYAKETLAS	300		
Db	241	tgnslvseensvatatnggalsgdavtisngsvtfsgnqavangaiyaklilas	300		
OY	301	GGGGISFSSNNIVGTTAGNGAISIIAAGBCSLAAGDITFNGNAIVATTPQTRKNS	360		
Db	301	ggggisfsmniygttagnngaisiiaagecslsaaagdltfngnaivattpqtkrns	360		
OY	361	IDIGSTAKITNLRAISGHSFEFDPITANTADSTDTLNKADAGNSTDYSGSIVSGE	420		
Db	361	idigstakicnlraissnifdydpitancaadstclnkaadagnstdysgsivsgge	420		
OY	421	KLSEDEAKVADNLTSLTKOPVTLTAGNLVLRKVTLDTKFTQTAGSSVIMDACTTLKAS	480		
Db	421	klseadeakvadnltsltkgvtltagnlvlkrgvltldtkftqtagssvimdagtlkas	480		
OY	481	TEETVTLGLSLIPVDSLGEGRKVYIAASAASKNALSPILLDNOGAYENHDLGKQDF	540		
Db	481	teetvltglslipvdslggrkvvyaasaasknvalsgp11lldnqgayenhdykqdf	540		
OY	541	SFVQSLAGTATTDDVAVPTVATPTHYGQGTWGMWVDFASTPKTKATLAWNTWGY	600		
Db	541	sfqslagattddvavptvatpthygqgtwgmwvddastpktkatlaawntlgy	600		
OY	601	LPNRQGLVPLNSLWGSFSDIQAIGVIERSAITLTCSDRGFMAAGVANFLDKDKGEKR	660		
Db	601	lpmrgqlvplnslwgsfadiqaiqviersaltlcsdrgfmaagvanflddkkgekr	660		
OY	661	KYRRKSGVAIGGAOTCSNMLISFAFCQJFGSDKQPLVAKNHTDITAGAFYIQHTIECS	720		
Db	661	kyrrksgyvaiggaotcsenlissafatcqlifgsdkdlvaknhtdyagaftyiqhtecs	720		
OY	721	GFIGCLDLKLPGSMSHRPVLEGOALAVSHVSNDLKTXYTAYPEVKGSGNNAFNMUGAS	780		
Db	721	gfignclldkpgsmshrpvlsegqlavshvsndlktktytpevksgnnafrmugas	780		
OY	781	SHSYPEYLHCFDITYAPYIKLNLTYIIRQDSFSEKGTGERSFSDSNLFLSLPIGVKPEKFS	840		
Db	781	shsypeylhcfdyapyiklnltyirgdsfsekgtegrsfdsdnlfslpivgkfkfs	840		
OY	841	DCNDFSVDLTLVYPDLIRNDPKCTRALVYSGASWETVANNLAQALQVAGSITFASPM	900		
Db	841	dcndfsvdltlvypdlirndpkctalvisgaswecyannlarqalqvragshyafspm	900		
OY	901	FEVLAGOFVFEVRGS 914			
Db	901	fevlagofvfevrgs 914			

RESULT 4  
 ID AAY90238 standard: Protein; 885 AA.  
 XX AAY90238;  
 XX AAY90238;

DT	29-AUG-2000 (first entry)	
XX		
DE	Mature Chlamydia antigen CPN100635.	
XX		
KW	Chlamydia antigen; diagnosis; infection; community acquired pneumonia;	
KW	therapy; upper respiratory tract disease; bronchitis; sinusitis;	
KW	asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma.	
XX		
OS	Chlamydia pneumoniae.	
XX		
PN	W0200032794-A2.	
PD	08-JUN-2000.	
XX		
PF	01-DEC-1999; 99NO-CA01147.	
XX		
PR	01-DEC-1998; 98US-0110339.	
PR	01-DEC-1998; 98US-0110340.	
PR	01-DEC-1998; 98US-0110427.	
PR	01-DEC-1998; 98US-0110428.	
PR	01-DEC-1998; 98US-0110438.	
XX		
PA	(CONN-) CONNADGHT LAB LTD.	
PI	Murdin AD, Oomen RP, Wang J;	
XX		
DR	WPI: 2000-412339/35.	
XX		
DR	N-PSDB: AAA30849, AAA30850.	
XX		
PT	Nucleic acids encoding polypeptide antigens from Chlamydia useful for	
PT	preventing, diagnosing and treating diseases such as community acquired	
PT	pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset	
PT	asthma -	
XX		
PS	Claim 16; Fig 3; 174pp: English.	
XX		
CC	This sequence is a Chlamydia antigen of the invention, designated	
CC	CPN100635. The nucleic acids (and their complementary sequences) may be	
CC	used as diagnostic agents for detecting the presence of nucleic acids	
CC	encoding Chlamydia antigens in samples according to standard methods,	
CC	and therefore, for diagnosing Chlamydia infections. For example, they may	
CC	be used as primers and probes for diagnostic polymerase chain reaction	
CC	(PCR) assays. Antisense sequences may be used to down regulate	
CC	expression of the proteins and may be used to treat infections. The	
CC	nucleic acids may also be used to produce the protein antigens they	
CC	encode according to standard recombinant DNA methodologies. The	
CC	proteins may then be used as antigens for the production of antibodies	
CC	(i.e. as vaccines) for preventing infection by Chlamydia. The	
CC	antibodies may also be used as diagnostic reagents for detecting	
CC	infections. Chlamydia is a pathogen implicated in the development of	
CC	(for example) community acquired pneumonia, upper respiratory tract	
CC	disease (especially bronchitis and sinusitis, asthmatic bronchitis,	
CC	adult-onset asthma and acute exacerbations of asthma in adults.	
XX		
SQ	Sequence 885 AA;	
XX		
Query Match	94.0%;	Score 4508; DB 21; Length 885;
Best Local Similarity	98.8%;	Pred. No. 4, 6e-288;
Matches 874; Conservative	1; Mismatches 10; Indels 0; Gaps 0;	
OY	44	GTYPRKNTTGGIDYTLTGDTLQNLGDSAAALTKGCFSDTTESLSFAGKGYSLSPFLNKSS 103
Db	1	gtyprknttgidytltgdtlqnlgsdaalatkcfdsdtteslsfagkyyslspflnkss 60
OY	104	AEGAALSVTTDKNLSLTGFSSTFLAAPSSTVITTPSGKAVKCGGDLTFDNNGTILRKOD 163
Db	61	aegaalsvttcknlsiltgfsstflaapsvlttppsgkavkcggdltfdnngtllfkkg 120
OY	164	YCEENGAISTRKLSLKNKSGSIFEGNKSATGKGAICATGTYDITNTWATPLFSNN 223
Db	121	yceengaislknlsknstgssifegnkssatgkkgacatgtydltntwaptlfsnn 180

CC encoding Chlamydia antigens in samples according to standard methods,  
 CC and therefore, for diagnosing Chlamydia infections. For example, they may  
 CC be used as primers and probes for diagnostic polymerase chain reaction  
 CC (PCR) assays. Antisense sequences may be used to down regulate  
 CC expression of the proteins and may be used to treat infections. The  
 CC nucleic acids may also be used to produce the protein antigens they  
 CC encode according to standard recombinant DNA methodologies. The  
 CC proteins may then be used as antigens for the production of antibodies  
 CC (i.e. as vaccines) for preventing infection by Chlamydia. The  
 CC antibodies may also be used as diagnostic reagents for detecting  
 CC infections. Chlamydia is a pathogen implicated in the development of  
 CC (for example) community acquired pneumonia, upper respiratory tract  
 CC disease (especially bronchitis and sinusitis, asthmatic bronchitis,  
 CC adult-onset asthma and acute exacerbations of asthma in adults.

XX Sequence 928 AA;

Query Match 98.6%; Score 4730; DB 21; Length 928;  
 Best Local Similarity 98.8%; Pred. No. 1.3e-302;  
 Matches 917; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 1 MKSQPSWLVLSSTLACFTSCSTVFPAATANTGSPDSFSGSTNTGTYTPKNTTGTIDYILT 60  
 DB 1 mksqpswlvlsstlacftscstvfpaatantgspdsfsgstntgtytpknttgtyilt 60  
 OY 61 GDITQNGDSALTKGCFSTTESLFAKGYSLSFLINIKSSAGALSVTTDNLSIT 120  
 DB 61 gditqngdsalatkcgfstteslfaagysslsfliniksaaagalsvttndnlsit 120  
 OY 121 GFSSLTFLAASSVITTPSGKAVKCGDLTFDNNGTILFKODYCENGAISTPKNSLK 180  
 DB 121 gfssltflaassvittpsgkavkcgdltfdnngtillfkodyceengastpknsllk 180  
 OY 181 NSTGSISFEGKSSATGKGAICATGTVDTNTNAPTPLFSNNIAEAAGAINSTGNCCTI 240  
 DB 181 nstgsisfegkssatgkgaicatgtvdtntnaptplfsnniaeaagainstgnccti 240  
 OY 241 TGNSTIVSESVATPATNGALSGDADVTISGNOSVFSGNOAVANGAITYAKKLTLAS 300  
 DB 241 tgnstivsestvapatngalsgdadvitlsgnosvftsgnvaangaiyakklilas 300  
 OY 301 GGGGGSISFNNIVGTTAGNGAISILAAECSSAEAGDITFNGNAIVATTPQTKRNS 360  
 DB 301 gggggssisfnnivgttagngaisilaaecssaeagditfngnaivattpqtkrns 360  
 OY 361 IDIGSTAKITMLRAISGISIFPYDPTANTAADSTDTNLNKADAGNSTDYSGSIVSGE 420  
 DB 361 idigstakitmlraisgisifpydptantaaadstdtlnlnkadagnstdysgsivsg 420  
 OY 421 KLSDEKAVANLSTLKOPTITAGNLVLRGVLTDRKGFQRTAGSSVINDAGTTLKAS 480  
 DB 421 klsedekavanlstlkkoptitagnlvlrgvltdrkqfqtqagssvindaagttlkas 480  
 OY 481 TEETVLTGLSIPVDSLGKRVIAAASAKNVALSGPILLLDNGNAYENHDLKTDQDF 540  
 DB 481 teetvltglsipvdslgkrviaaasaknvalsgpilllldngnayenhdlgktqdf 540  
 OY 541 SFVOLALGATTTDDPAVPTVATPHKGYGTGWTGMTWDDTASTPKTKATLAWTNTGY 600  
 DB 541 sfvolalगतttddpavptvatphkgygtgwtgmtwddtastpkktactlawntgy 600  
 OY 601 LPNBERGPIVPSNLMSFSDIOALOGVTERSAITLCSDRFMAAGVNFIDKDKKGR 660  
 DB 601 lpnbergpivpnsnlmsfSDIOALOGVTERSAITLCSDRFMAAGVNFIDKDKKGR 660  
 OY 661 KYRRKSGGYAIGAAQCSSEMLISFAFCQLFGSKDFIVANKHDTYAGAYIOHITCS 720  
 DB 661 kyrrksGGYAIgaaqcssemlisfaFCQLfgskdfivankhdtYagayiohitcs 720  
 OY 721 GFICGLLDKLPGSWSHKPLVIEGOLAVSHVSNLKTXTATPEVKGSGNNAFNMILGAS 780  
 DB 721 gficgllDkLPgswshkplviegolavshvsnlktxtatpevkgsgnnaFnmilgas 780

OY 781 SHSYPEYLHCEFTYAPYIKLANLTYIRODSFSEKTEGHSFDDSNLFNLPIGVKPERFS 840  
 DB 781 shsypeylhceftypyklnlntyirodsfsekteghsfddsnlnfslpvgvkefs 840  
 OY 841 DCNDRSYDLTLYVDLIRNDPKCTTALVIGASVETANNIARALOVRAGSHTAFSPM 900  
 DB 841 dcnDRsydlTlyvdlirndpkcttalvigsavetanniaralovragshyafspm 900  
 OY 901 FEVLGQFVEVRGSSRINVDGKQFOF 928  
 DB 901 fevlqfvevrgssrinyvdlgkqf 928

# RESULT 3

AAW88429 standard; Protein: 914 AA.  
 ID AAW88429  
 AC AAW88429;  
 AC AAW88429;  
 DT 26-APR-1999 (first entry)  
 DE Chlamydia pneumoniae surface exposed protein Omp5.  
 DE Omp5; outer membrane protein 5; surface exposed protein; antigen;  
 KW infection; diagnosis; vaccine; atherosclerosis; asthma.  
 KW Chlamydia pneumoniae.  
 OS Chlamydia pneumoniae.  
 PN W09858953-A2.  
 PN 30-DEC-1998.  
 PD 19-JUN-1998; 98WO-DK00265.  
 PE 23-JUN-1997; 97DK-0000744.  
 PR (BIRKELUND S.  
 PA (CHRIL/) CHRISTIANSEN G.  
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;  
 PI Mygind P;  
 DR WPI: 1999-105610/09.  
 DR N-PSDB: AAX06828.  
 XX Species-specific test for identifying mammals infected with  
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for  
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding  
 PT these proteins  
 PS Disclosure: Page 78-80; 115pp; English.  
 PS This polypeptide is described as a subsequence of the claimed  
 CC novel surface exposed protein Omp5 (see AAW88418) of Chlamydia  
 CC pneumoniae, a human respiratory pathogen. The invention provides  
 CC a new species specific test for identifying mammals (including  
 CC humans) infected with C. pneumoniae. The test comprises detecting  
 CC antibodies specific for surface exposed proteins Omp5 (see  
 CC AAW88417-28) or detecting nucleic acid fragments encoding them (see  
 CC AAW06816-27), especially by PCR. The proteins are also used in the  
 CC diagnosis of C. pneumoniae infection in mammals. The nucleic acids  
 CC and proteins can also be used in the immunization of mammals, the  
 CC nucleic acids being particularly useful as DNA vaccines for  
 CC effecting in vivo expression of antigens. The vaccines may also  
 CC prevent atherosclerosis and bronchial asthma, which are possibly  
 CC associated with C. pneumoniae.  
 XX Sequence 914 AA;

Query Match 98.3%; Score 4714; DB 20; Length 914;  
 Best Local Similarity 99.9%; Pred. No. 1.4e-301;



PT outer membrane proteins of *C. pneumoniae* or nucleic acids encoding  
 PT these proteins  
 XX  
 PS Claim 7; Page 43-45; 115pp; English.  
 CC This polypeptide comprises the novel 97.2 kDa surface exposed  
 CC protein Omp5 of the human respiratory pathogen *Chlamydia*  
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see  
 CC AX006817) isolated from a *C. pneumoniae* expression library. The  
 CC invention provides 12 novel surface exposed proteins, Omp4-Omp15  
 CC (see AW8417-28), and nucleic acid sequences encoding them (see  
 CC AX006816-27). A new species specific test is claimed that is used  
 CC to identify mammals (including humans) infected with *Chlamydia*  
 CC pneumoniae. The test comprises detecting antibodies specific for  
 CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer  
 CC membrane proteins, especially by PCR. The proteins are also used  
 CC in the diagnosis of *C. pneumoniae* infection in mammals. The  
 CC nucleic acids and proteins can also be used in the immunization of  
 CC mammals, the nucleic acids being particularly useful as DNA  
 CC vaccines for effecting in vivo expression of antigens. The  
 CC vaccines may also prevent atherosclerosis and bronchial asthma,  
 CC which are possibly associated with *C. pneumoniae*.  
 CC  
 XX  
 SQ Sequence 928 AA;

Query Match 100.0%; Score 4795; DB 20; Length 928;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSQFSLVLSSTLACFTSCSTVFAATAENIGPSDFSGSTNTGTYPKMTTGIDYTLT 60  
 DB 1 mksqfslvlstlactfscstvfataenlgspsdfsgstntgtypkmtttgidytlc 60  
 QY 61 GDTTLQMLGSAALTKCFEDTRESLSFAGKGYSLSLINKSSAKGALSVYTDKNLSLT 120  
 DB 61 gdtltqlglsaalkcfedteslsfagkgysslslnkssagaalsvtldknlslt 120  
 QY 121 GFSLFLFAPSSVITTPSGKGAIVGCGDLTFPDNNGTILRKQDVCENGGASITRNLSLK 180  
 DB 121 gfslflfapssvlttppsgkgaivgcgdltfpdngtllrkqdcenngaisltnlslk 180  
 QY 181 NSTGSISEFGNKSATGKKGATCATGTVDITNTAPTLFSNNIAEAGALNSTGNCIT 240  
 DB 181 nstgsisefgnksatgkkgatcatgtvdtntaptlfsnniaeaagalnstgncit 240  
 QY 241 TGNTSLVSESNVTATAGNGALSGDADVTISGNQSVTFSGNOAVANGALYAKKLTLAS 300  
 DB 241 tgnstslvseesnvtatagngalsgdadvitsgnqsvtfsgnqavangalylakkltlas 300  
 QY 301 GGGGSISESNNIYOGTATAGNGALSTLAAGECSISAFAGDITFNGNIVATTPTTKRNS 360  
 DB 301 ggggsisefsnniyogtataagngalstlaagecsisafagditfngnivattptrkns 360  
 QY 361 IDIGSTAKITNLRAISGHSIFEFDPITANTADSTDTLNLKADAGNSTDYSGSIVSEGE 420  
 DB 361 idigstakitnlraishsifefdpitantadstdtlnlkadagnstdysgsivsege 420  
 QY 421 KLSDEKAVADNLTSTLKQPVYTLTAGMLVLKRGVYTLDTKFTQTAGSSVIMDAGTTLKAS 480  
 DB 421 klsedekavadnltstlkqpytltagmlvlkrgvtyltdtkftqtagssvimdagttlkas 480  
 QY 481 TEEVTLTGLSPVDSLGEKKVYTAAGSKNVALSGPILLDNQAGKAYNHHDGKQDF 540  
 DB 481 teevtltglsplvdslegkkyvtaagsknvalsgpilldnqagaynhhdgkqdf 540  
 QY 541 SPVQLSALGATTTDPAVPFVAIPTHYGYQGTWGMWVDRTASTPKTKATLAWNTGY 600  
 DB 541 spvqlsalgatttdpavpfvaptthygyqgtwgmwvdrtastpktkatlawnnty 600  
 QY 601 LPPPERGGLVPVPSIMSFSDIOALGVIRRSALTLCSDGFMAGVANGVNLDDKKEKR 660  
 DB 601 lpppergglvpvpsimsfedioalgvirrsaltlcSDGFMAGVANGVNLDDKKEKR 660  
 DB 601 lpppergglvpvpsimsfedioalgvirrsaltlcSDGFMAGVANGVNLDDKKEKR 660

QY 661 KYRHKSGYALIGAAORCSENLISFAFCOLFGSDKDFLVAKNHTDTYAGAFYIOHITECS 720  
 DB 661 kyrrksqyaligaaorcsenlisfarcolfgsdkdfvlvaknhtdtyagafyiohitecs 720  
 QY 721 GFTEGLDLKLPGSWSHKLPLEGOLAYSHVSNDLKRTYTAPEYKSGWGNNAFMNLGAS 780  
 DB 721 gfteglldklpgswshklplegolaishvsnldkrtytapeykgswgnnafmnlgas 780  
 QY 781 SHSTPEFLHCEDTYAPYIKMLTYIRODSFSEKTEBGRSPDSNLFULSLPIGVKFEKFS 840  
 DB 781 shstpeylhcedtyapyikmltyirodsfsektebgrspdsnlfulslpivgkfkfs 840  
 QY 841 DCNDFSYDLTSLVPDLIRNDPKCTALVTSAGASWENYANNLAQALQVRAGSHYAFSPM 900  
 DB 841 dcndfsydltslvpdlirndpkctalvtisgawenyannlaqalqvragshtyafspm 900  
 QY 901 FEVLGQFVFEVGRSGSRITVNDLGKRFQF 928  
 DB 901 fevlgqfvfevgrsgsrityvndlgkrfqf 928

RESULT 2  
 AAY90237  
 ID AAY90237 standard; Protein; 928 AA.  
 XX  
 AC AAY90237;  
 XX  
 DT 29-AUG-2000 (first entry)  
 XX  
 XX Chlamydia antigen CPN100635.  
 DE  
 XX Chlamydia antigen: diagnosis; infection; community acquired pneumonia;  
 KW therapy; upper respiratory tract disease; bronchitis; sinusitis;  
 KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma.  
 XX Chlamydia pneumoniae.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..43  
 FT /note= "signal peptide"  
 FT Protein 44..928  
 FT /note= "mature CPN100635"

PN MO200032794-A2.  
 PD 08-JUN-2000.  
 PD 01-DEC-1999; 99MO-CA01147.  
 PF 01-DEC-1998; 98US-0110339.  
 PR 01-DEC-1998; 98US-0110340.  
 PR 01-DEC-1998; 98US-0110427.  
 PR 01-DEC-1998; 98US-0110428.  
 PR 01-DEC-1998; 98US-0110438.  
 XX  
 PA (CONN-) CONNACHT LAB LTD.  
 XX  
 PI Murlin AD, Oomen RP, Wang J;  
 XX  
 DR WPI; 2000-412339/35.  
 DR N-PSDB; AAA30849, AAA30850.  
 XX

PT Nucleic acids encoding polypeptide antigens from *Chlamydia* useful for  
 PT preventing, diagnosing and treating diseases such as community acquired  
 PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset  
 PT asthma -  
 PS  
 XX  
 PS Claim 16; Fig 3; 174pp; English.

CC This sequence is a *Chlamydia* antigen of the invention, designated  
 CC CPN100635. The nucleic acids (and their complementary sequences) may be  
 CC used as diagnostic agents for detecting the presence of nucleic acids

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein. - protein search, using sw model

Run on: October 2, 2001, 03:26:04 : Search time 97.54 Seconds  
(without alignments)  
576.779 Million cell updates/sec

Title: US-09-446-677b-4  
Perfect score: 4795  
Sequence: 1 MKSQFSLVLTSLACFTSC.....FEVGRSRIYVDLGGKQF 928

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq\_0601.\*  
1: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseqp/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4795	100.0	928	20	AAW88418
2	4730	98.6	928	21	AAV90237
3	4714	98.3	914	20	AAW88429
4	4508	94.0	885	21	AAV90238
5	3275	68.3	643	20	AAV35056
6	2237	46.7	918	21	AAV69369
7	2222	46.3	918	20	AAW88422
8	2069	43.1	928	20	AAW88421
9	2058	42.9	928	21	AAV94327
10	2030	42.3	930	21	AAV90240
11	2028	42.3	930	20	AAV35052

12	2024	42.2	930	20	AAW88424	Chlamydia pneumoni
13	1935.5	40.4	945	21	AAV69368	Amino acid sequenc
14	1931	40.3	928	21	AAW88423	Chlamydia pneumoni
15	1920	40.0	928	21	AAV90239	Chlamydia antigen
16	1908.5	39.8	945	20	AAW88428	Chlamydia pneumoni
17	1881.5	39.2	927	20	AAV35054	Chlamydia pneumoni
18	1855	38.7	936	21	AAV99842	Chlamydia pneumoni
19	1827.5	38.1	925	21	AAV99843	Chlamydia pneumoni
20	1786	37.2	948	20	AAV35060	Chlamydia pneumoni
21	1785	37.2	928	20	AAW88417	Chlamydia pneumoni
22	1785	37.2	928	21	AAV90236	Chlamydia antigen
23	1403.5	29.3	841	21	AAV92818	C. pneumoniae CPN1
24	1400.5	29.2	841	20	AAW88420	Chlamydia pneumoni
25	1398	29.2	922	21	AAV95548	Chlamydia pneumoni
26	1396.5	29.1	507	20	AAV34614	Chlamydia pneumoni
27	1393	29.1	922	20	AAW88419	Chlamydia pneumoni
28	1386	28.9	922	20	AAV34597	Chlamydia pneumoni
29	1341.5	28.0	973	21	AAV96274	Chlamydia pneumoni
30	1267	26.4	597	20	AAV34611	Chlamydia pneumoni
31	1254.5	26.2	671	20	AAV35050	Chlamydia pneumoni
32	1116	23.3	450	20	AAV34617	Chlamydia pneumoni
33	1102.5	23.0	1013	20	AAV16738	C. trachomatis F s
34	1101	23.0	1012	20	AAV16735	C. trachomatis LGV
35	1100.5	23.0	1013	20	AAV16737	C. trachomatis B s
36	1094	22.8	982	21	AAV16733	C. trachomatis pmp
37	1089	22.7	1006	21	AAV16739	C. trachomatis pmp
38	1041	21.7	1132	20	AAV35048	Chlamydia pneumoni
39	902	18.8	610	20	AAW88431	Chlamydia pneumoni
40	885	18.5	880	21	AAV16732	C. trachomatis pmp
41	876	18.3	866	21	AAV16738	C. trachomatis pmp
42	816.5	17.0	494	20	AAV34615	Chlamydia pneumoni
43	814	17.0	427	20	AAV34613	Chlamydia pneumoni
44	812.5	16.9	483	20	AAV34609	Chlamydia pneumoni
45	777	16.2	530	20	AAV35064	Chlamydia pneumoni

## ALIGNMENTS

RESULT 1	
AAW88418	standard; Protein: 928 AA.
XX	
AC	AAW88418;
XX	
DT	26-APR-1999 (first entry)
XX	
DE	Chlamydia pneumoniae surface exposed protein Omp5.
XX	
KW	Omp5; outer membrane protein 5; surface exposed protein; antigen;
KM	infection; diagnosis; vaccine; atherosclerosis; asthma.
XX	
OS	Chlamydia pneumoniae.
XX	
PN	W09858953-A2.
XX	
PD	30-DEC-1998.
XX	
PF	19-JUN-1998; 98WO-DK00266.
XX	
PR	23-JUN-1997; 97DK-0000744.
XX	
PA	(BIRK/) BIRKELUND S.
XX	(CHR/) CHRISTIANSEN G.
PI	Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI	Mygind P;
DR	WPT, 1999-105610/09.
XX	N-PSDB; AAX06817.
PT	Species-specific test for identifying mammals infected with
	Chlamydia pneumoniae - comprises detecting antibodies specific for